

# ONLINE SEARCH REQUEST FORM

USER

Specton

U.S. SERIAL NUMBER

223263

ART UNIT

1812

PHONE

308-1793

DATE \_\_\_\_\_

4/10/95

Please give a detailed statement of requirements. Describe as specifically as possible the subject matter to be searched. Define any terms that may have special meaning. Give examples or relevant citations, authors, or keywords, if known.

You may include a copy of the broadest and or relevant claim(s).

Please search ~~89~~<sup>10</sup> 142

**STAFF USE ONLY**

**COMPLETED**

SEARCHER

## ONLINE TIME

(in minutes)

NO. OF DATABASES

TOTAL TIME

## SYSTEMS

CAS ONLINE

DARC/QUESTEL

## DIALOG

SDC

OTHER



3.	P50301	Monkey erythropoietin.	192	20	43	6.71	0
4.	P50343	EPO encoded by clone lambda-H	193	20	44	6.71	0
5.	R28150	Sugar beet chitinase 1.	439	20	40	6.71	0
6.	R20229	DOG-1.	350	19	60	6.29	0
7.	R45155	**** 5 standard deviations above mean ****	876	18	60	5.87	0
8.	R28348	B. atterothomophilus DNA poly	877	18	60	5.87	0
9.	R14854	Bacillus caldophilus DNA poly	900	18	64	5.87	0
10.	R12538	Protein associated with bioch	119	17	30	5.45	0
11.	R07674	SMOC-87 intestinal mucin cDNA	119	17	30	5.45	0
12.	R33420	Human IL-8 receptor from clon	335	17	39	5.45	0
13.	R38273	Sequence in a low affinity re	350	17	39	5.45	0
14.	R14838	Protein deduced from human G1	405	17	71	5.45	0
15.	R23970	MPV env protein with growth	635	17	33	5.45	0
16.	R20637	Human Cytomegalovirus antigen	1048	17	64	5.45	0
17.	R34257	Anti-HIV gp120 immunoglobulin	106	16	25	5.03	0
18.	R10532	Prod. of DNA of PMG08 used to	156	16	36	5.03	0
19.	R24137	Fish growth hormone.	188	16	39	5.03	0
20.	R07353	Fish growth hormone polypepti	188	16	39	5.03	0
21.	P90732	Fish growth hormone.	200	16	41	5.03	0
22.	P93638	Fish growth hormone.	204	16	43	5.03	0
23.	R43684	Lipopolysaccharide induced pr	311	16	48	5.03	0
24.	R12345	Toxoplasma gondii P66 antigen	332	16	45	5.03	0
25.	R12352	Toxoplasma gondii P66 antigen	428	16	44	5.03	0
26.	R12098	Haem 84-1 portion of H.confor	516	16	53	5.03	0
27.	P81865	Sequence encoded by LAV MA L	889	16	72	5.03	0
28.	R04232	Rickettsia rickettsii p120 an	1267	16	66	5.03	0
29.	P81184	Sequence encoded by the 2nd r	1534	16	59	5.03	0
30.	R04032	Full length T4 encoded by pla	2037	16	61	5.03	0
31.	R12327	**** 4 standard deviations above mean ****	128	15	27	4.61	0
32.	R54979	Light (kappa) chain variable	128	15	28	4.61	0
33.	R54979	Brassica seed acyl carrier pr	134	15	28	4.61	0
34.	P81139	Sequence of plant acyl carrie	134	15	28	4.61	0
35.	R23593	Recombinant hematoipoietic mol	156	15	38	4.61	0
36.	P70398	Sequence of human erythropoie	166	15	38	4.61	0
37.	P50299	Human recombinant erythropoie	167	15	38	4.61	0
38.	P50298	Human recombinant erythropoie	167	15	38	4.61	0
39.	R07665	Cysteine-added variant of ery	167	15	38	4.61	0
40.	R31608	Homologous to chicken nov gen	205	15	26	4.61	0
41.	R12798	OV3 light and heavy chains.	241	15	31	4.61	0
42.	R43576	Bovine adrenocorticotrophic ho	266	15	42	4.61	0
43.	R23598	Recombinant hematoipoietic mol	303	15	63	4.61	0
44.	P81843	Recombinant hematoipoietic mol	322	15	64	4.61	0
45.	R37486	Sequence of alkaline phosphat	409	15	67	4.61	0
		Uojoba fatty acyl reductase e	524	15	48	4.61	0

1. US-08-223-263-1 (1-353)  
R29164 PRP3.

ID R29164 standard; Protein; 121 AA.  
AC R29164;  
DT 21-APR-1993 (first entry)  
DE PRP3.  
KW Proline-rich protein; PRP1; PRP2; PRP3; PRP4; PRP3L2; pistill;  
flower; sterile; plant; hydroxyproline; glycoprotein; gum.  
OS Nicotiana glauca.  
FT Key Location/Qualifiers  
FT 23..24  
FT /note= "Ser-Pro repeat unit"

FT	Region	27..28
FT	/note=	"Ser-Pro repeat unit"
FT	Region	31..32
FT	/note=	"Ser-Pro repeat unit"
FT	Region	35..38
FT	/note=	"Ser-(Pro)3 repeat unit"
FT	Region	44..48
FT	/note=	"Ser-(Pro)4 repeat unit"
FT	Region	51..55
FT	/note=	"Ser-(Pro)4 repeat unit"
FT	Region	56..60
FT	/note=	"Ser-(Pro)4 repeat unit"
FT	Region	61..62
FT	/note=	"Ser-Pro repeat unit"
FT	Region	63..67
FT	/note=	"Ser-(Pro)4 repeat unit"
FT	Region	68..69
FT	/note=	"Ser-Pro repeat unit"
FT	Region	70..74
FT	/note=	"Ser-(Pro)4 repeat unit"
FT	Region	75..76
FT	/note=	"Ser-Pro repeat unit"
FT	Region	77..81
FT	/note=	"Ser-(Pro)4 repeat unit"
FT	Region	82..84
FT	/note=	"Ser-(Pro)2 repeat unit"
FT	Region	91..92
FT	/note=	"Ser-Pro repeat unit"
FT	Region	93..94
FT	/note=	"Ser-Pro repeat unit"
FT	Region	95..99
FT	/note=	"Ser-Pro repeat unit"
FT	Region	100..104
FT	/note=	"Ser-Pro repeat unit"
FT	Region	105..109
FT	/note=	"Ser-Pro repeat unit"
FT	Region	110..114
FT	/note=	"Ser-Pro repeat unit"
FT	Region	115..119
FT	/note=	"Ser-Pro repeat unit"
FT	Region	120..124
FT	/note=	"Ser-Pro repeat unit"
FT	Region	125..129
FT	/note=	"Ser-Pro repeat unit"
FT	Region	130..134
FT	/note=	"Ser-Pro repeat unit"
FT	Region	135..139
FT	/note=	"Ser-Pro repeat unit"
FT	Region	140..144
FT	/note=	"Ser-Pro repeat unit"
FT	Region	145..149
FT	/note=	"Ser-Pro repeat unit"
FT	Region	150..154
FT	/note=	"Ser-Pro repeat unit"
FT	Region	155..159
FT	/note=	"Ser-Pro repeat unit"
FT	Region	160..164
FT	/note=	"Ser-Pro repeat unit"
FT	Region	165..169
FT	/note=	"Ser-Pro repeat unit"
FT	Region	170..174
FT	/note=	"Ser-Pro repeat unit"
FT	Region	175..179
FT	/note=	"Ser-Pro repeat unit"
FT	Region	180..184
FT	/note=	"Ser-Pro repeat unit"
FT	Region	185..189
FT	/note=	"Ser-Pro repeat unit"
FT	Region	190..194
FT	/note=	"Ser-Pro repeat unit"
FT	Region	195..199
FT	/note=	"Ser-Pro repeat unit"
FT	Region	200..204
FT	/note=	"Ser-Pro repeat unit"
FT	Region	205..209
FT	/note=	"Ser-Pro repeat unit"
FT	Region	210..214
FT	/note=	"Ser-Pro repeat unit"
FT	Region	215..219
FT	/note=	"Ser-Pro repeat unit"
FT	Region	220..224
FT	/note=	"Ser-Pro repeat unit"
FT	Region	225..229
FT	/note=	"Ser-Pro repeat unit"
FT	Region	230..234
FT	/note=	"Ser-Pro repeat unit"
FT	Region	235..239
FT	/note=	"Ser-Pro repeat unit"
FT	Region	240..244
FT	/note=	"Ser-Pro repeat unit"
FT	Region	245..249
FT	/note=	"Ser-Pro repeat unit"
FT	Region	250..254
FT	/note=	"Ser-Pro repeat unit"
FT	Region	255..259
FT	/note=	"Ser-Pro repeat unit"
FT	Region	260..264
FT	/note=	"Ser-Pro repeat unit"
FT	Region	265..269
FT	/note=	"Ser-Pro repeat unit"
FT	Region	270..274
FT	/note=	"Ser-Pro repeat unit"
FT	Region	275..279
FT	/note=	"Ser-Pro repeat unit"
FT	Region	280..284
FT	/note=	"Ser-Pro repeat unit"
FT	Region	285..289
FT	/note=	"Ser-Pro repeat unit"
FT	Region	290..294
FT	/note=	"Ser-Pro repeat unit"
FT	Region	295..299
FT	/note=	"Ser-Pro repeat unit"
FT	Region	300..304
FT	/note=	"Ser-Pro repeat unit"
FT	Region	305..309
FT	/note=	"Ser-Pro repeat unit"
FT	Region	310..314
FT	/note=	"Ser-Pro repeat unit"
FT	Region	315..319
FT	/note=	"Ser-Pro repeat unit"
FT	Region	320..324
FT	/note=	"Ser-Pro repeat unit"
FT	Region	325..329
FT	/note=	"Ser-Pro repeat unit"
FT	Region	330..334
FT	/note=	"Ser-Pro repeat unit"
FT	Region	335..339
FT	/note=	"Ser-Pro repeat unit"
FT	Region	340..344
FT	/note=	"Ser-Pro repeat unit"
FT	Region	345..349
FT	/note=	"Ser-Pro repeat unit"
FT	Region	350..354
FT	/note=	"Ser-Pro repeat unit"
FT	Region	355..359
FT	/note=	"Ser-Pro repeat unit"
FT	Region	360..364
FT	/note=	"Ser-Pro repeat unit"
FT	Region	365..369
FT	/note=	"Ser-Pro repeat unit"
FT	Region	370..374
FT	/note=	"Ser-Pro repeat unit"
FT	Region	375..379
FT	/note=	"Ser-Pro repeat unit"
FT	Region	380..384
FT	/note=	"Ser-Pro repeat unit"
FT	Region	385..389
FT	/note=	"Ser-Pro repeat unit"
FT	Region	390..394
FT	/note=	"Ser-Pro repeat unit"
FT	Region	395..399
FT	/note=	"Ser-Pro repeat unit"
FT	Region	400..404
FT	/note=	"Ser-Pro repeat unit"
FT	Region	405..409
FT	/note=	"Ser-Pro repeat unit"
FT	Region	410..414
FT	/note=	"Ser-Pro repeat unit"
FT	Region	415..419
FT	/note=	"Ser-Pro repeat unit"
FT	Region	420..424
FT	/note=	"Ser-Pro repeat unit"
FT	Region	425..429
FT	/note=	"Ser-Pro repeat unit"
FT	Region	430..434
FT	/note=	"Ser-Pro repeat unit"
FT	Region	435..439
FT	/note=	"Ser-Pro repeat unit"
FT	Region	440..444
FT	/note=	"Ser-Pro repeat unit"
FT	Region	445..449
FT	/note=	"Ser-Pro repeat unit"
FT	Region	450..454
FT	/note=	"Ser-Pro repeat unit"
FT	Region	455..459
FT	/note=	"Ser-Pro repeat unit"
FT	Region	460..464
FT	/note=	"Ser-Pro repeat unit"
FT	Region	465..469
FT	/note=	"Ser-Pro repeat unit"
FT	Region	470..474
FT	/note=	"Ser-Pro repeat unit"
FT	Region	475..479
FT	/note=	"Ser-Pro repeat unit"
FT	Region	480..484
FT	/note=	"Ser-Pro repeat unit"
FT	Region	485..489
FT	/note=	"Ser-Pro repeat unit"
FT	Region	490..494
FT	/note=	"Ser-Pro repeat unit"
FT	Region	495..499
FT	/note=	"Ser-Pro repeat unit"
FT	Region	500..504
FT	/note=	"Ser-Pro repeat unit"
FT	Region	505..509
FT	/note=	"Ser-Pro repeat unit"
FT	Region	510..514
FT	/note=	"Ser-Pro repeat unit"
FT	Region	515..519
FT	/note=	"Ser-Pro repeat unit"
FT	Region	520..524
FT	/note=	"Ser-Pro repeat unit"
FT	Region	525..529
FT	/note=	"Ser-Pro repeat unit"
FT	Region	530..534
FT	/note=	"Ser-Pro repeat unit"
FT	Region	535..539
FT	/note=	"Ser-Pro repeat unit"
FT	Region	540..544
FT	/note=	"Ser-Pro repeat unit"
FT	Region	545..549
FT	/note=	"Ser-Pro repeat unit"
FT	Region	550..554
FT	/note=	"Ser-Pro repeat unit"
FT	Region	555..559
FT	/note=	"Ser-Pro repeat unit"
FT	Region	560..564
FT	/note=	"Ser-Pro repeat unit"
FT	Region	565..569
FT	/note=	"Ser-Pro repeat unit"
FT	Region	570..574
FT	/note=	"Ser-Pro repeat unit"
FT	Region	575..579
FT	/note=	"Ser-Pro repeat unit"
FT	Region	580..584
FT	/note=	"Ser-Pro repeat unit"
FT	Region	585..589
FT	/note=	"Ser-Pro repeat unit"
FT	Region	590..594
FT	/note=	"Ser-Pro repeat unit"
FT	Region	595..599
FT	/note=	"Ser-Pro repeat unit"
FT	Region	600..604
FT	/note=	"Ser-Pro repeat unit"
FT	Region	605..609
FT	/note=	"Ser-Pro repeat unit"
FT	Region	610..614
FT	/note=	"Ser-Pro repeat unit"
FT	Region	615..619
FT	/note=	"Ser-Pro repeat unit"
FT	Region	620..624
FT	/note=	"Ser-Pro repeat unit"
FT	Region	625..629
FT	/note=	"Ser-Pro repeat unit"
FT	Region	630..634
FT	/note=	"Ser-Pro repeat unit"
FT	Region	635..639
FT	/note=	"Ser-Pro repeat unit"
FT	Region	640..644
FT	/note=	"Ser-Pro repeat unit"
FT	Region	645..649
FT	/note=	"Ser-Pro repeat unit"
FT	Region	650..654
FT	/note=	"Ser-Pro repeat unit"
FT	Region	655..659
FT	/note=	"Ser-Pro repeat unit"
FT	Region	660..664
FT	/note=	"Ser-Pro repeat unit"
FT	Region	665..669
FT	/note=	"Ser-Pro repeat unit"
FT	Region	670..674
FT	/note=	"Ser-Pro repeat unit"
FT	Region	675..679
FT	/note=	"Ser-Pro repeat unit"
FT	Region	680..684
FT	/note=	"Ser-Pro repeat unit"
FT	Region	685..689
FT	/note=	"Ser-Pro repeat unit"
FT	Region	690..694
FT	/note=	"Ser-Pro repeat unit"
FT	Region	695..699
FT	/note=	"Ser-Pro repeat unit"
FT	Region	700..704
FT	/note=	"Ser-Pro repeat unit"
FT	Region	705..709
FT	/note=	"Ser-Pro repeat unit"
FT	Region	710..714
FT	/note=	"Ser-Pro repeat unit"
FT	Region	715..719
FT	/note=	"Ser-Pro repeat unit"
FT	Region	720..724
FT	/note=	"Ser-Pro repeat unit"
FT	Region	725..729
FT	/note=	"Ser-Pro repeat unit"
FT	Region	730..734
FT	/note=	"Ser-Pro repeat unit"
FT	Region	735..739
FT	/note=	"Ser-Pro repeat unit"
FT	Region	740..744
FT	/note=	"Ser-Pro repeat unit"
FT	Region	745..749
FT	/note=	"Ser-Pro repeat unit"
FT	Region	750..754
FT	/note=	"Ser-Pro repeat unit"
FT	Region	755..759
FT	/note=	"Ser-Pro repeat unit"
FT	Region	760..764
FT	/note=	"Ser-Pro repeat unit"
FT	Region	765..769
FT	/note=	"Ser-Pro repeat unit"
FT	Region	770..774
FT	/note=	"Ser-Pro repeat unit"
FT	Region	775..779
FT	/note=	"Ser-Pro repeat unit"
FT	Region	780..784
FT	/note=	"Ser-Pro repeat unit"
FT	Region	785..789
FT	/note=	"Ser-Pro repeat unit"
FT	Region	790..794



DT 01-JAN-1980 (first entry)  
DE Monkey erythropoietin.  
KW Erythropoietin; red blood cell; erythrocyte; anaemia; blood;  
OS disorder; ds.  
PN Monkey.  
W08502610-A.  
PD 20-JUN-1985.  
PF 11-DEC-1984; U002021.  
PR 13-DEC-1983; US-561024.  
PR 21-FEB-1984; US-582183.  
PR 28-SEP-1984; US-653841.  
PR 30-NOV-1984; US-675298.  
PA (KIRL-) KIRIN-AMGEN INC.  
DR WPI; 85-159229/26.  
DR N-PSDB; N50348.  
PT New polypeptide having properties of erythropoietin - is prepd.  
PT by cultivation of transformed eucaryotic or procaryotic host  
PS Disclosure; Page 38; 113pp; English.  
CC Monkey erythropoietin encoded by this sequence is essential for red  
CC blood cell formation and is used for the diagnosis and treatment of  
CC blood disorders such as anaemia. Large amounts of EPO may be obtained  
CC using recombinant DNA techniques in contrast to small amounts  
CC obtained from plasma and urine. See also N50345-47, N50349-50  
CC and P50298-P50300.  
SQ Sequence 192 AA;  
SQ 20 A; 13 R; 6 N; 6 D; 0 B; 6 C; 7 Q; 13 E; 0 Z; 12 G; 2 H;  
SQ 6 I; 30 L; 7 K; 4 M; 5 F; 11 P; 14 S; 9 T; 4 W; 4 Y; 13 V;  
Initial Score = 20 Optimized Score = 43 Significance = 6.71  
Residue Identity = 27% Matches = 54 Mismatches = 117  
Gaps = 27 Conservative Substitutions = 0  
X 10 20 30 40 50 60  
MELTELLVWMLL-TARLTSSP---APP--ACDLRVLSKLRLDSVHLSRLSQCEVHPLEPTVLLPAND  
IGVHCPCPAMLWLLSLVSLPLGLVPRGAPRLICDSVLERVYLLEAKENVTMGSCSINENITVDDTK  
X 10 20 30 40 50 60 70  
FSLGEKTMETKADILGAVTLLEGVMAARQGLPTCLSSLLGOL-----SG---QVRL--LGA-  
-LQSLIGTLPQGRKTAHKDPNALFTSFQHLRKGVRFLMVLVGGSTLCVRRAPPTTAVPSRST  
EALISPDASAPLRTITADTFCKLFRVSNFLRGK---LKIYTG-EACRRGDR  
X 130 140 150 160 170 180 190  
VNFYAKRREVAQAQVEVWOGIALISEAV--LRGQAVLVNSSQPFEPQLQHMKAISGLRSTLTLRAIGAQ  
X 150 160 170 180 190  
4. US-08-223-263-1 (1-353)  
P50343 EPO encoded by clone lambda-HEPOFL13.  
ID P50343 standard; Protein; 193 AA.  
AC P50343;  
DT 10-MAR-1992 (first entry)  
DE EPO encoded by clone lambda-HEPOFL13.  
KW Erythropoietin.  
OS Homo sapiens.

FH Key Location/Qualifiers  
FT Peptide 1..27  
FT /label= signal\_sequence  
FT Protein 27..193  
FT /label= mature\_EPO  
PN DK8406107-A.  
PD 16-AUG-1985.  
PF 27-DEC-1984; US-565627.  
PR 27-DEC-1983; US-566037.  
PR 04-DEC-1984; US-677813.  
PR 21-DEC-1984; NO-005186.  
PR 21-DEC-1984; ZA-010034.  
PR 22-JAN-1985; US-693258.  
PR 22-JAN-1985; US-688622.  
PA (GENE-) GENETICS INST.  
PA Kauffman RJ.  
PI WPI; 85-318061/51.  
DR N-PSDB; N50443.  
DT Vector system for introducing heterologous DNA into eukaryotic  
PT cells - comprises prod. gene and accessory DNA for enhanced  
PT expression of heterologous protein by the cells.  
PS Disclosure; Fig 13; 62pp; Danish.  
CC See also US4740461 88.04.26 (8819) (first major country equivalent).  
CC The sequence is encoded by clone lambda-HEPOFL13 which contains  
CC the gene encoding erythropoietin obt. from a human foetal DNA  
CC library. The gene can be expressed using a novel vector system  
CC disclosed in the specification.  
SQ Sequence 193 AA;  
SQ 20 A; 14 R; 6 N; 6 D; 0 B; 4 C; 8 Q; 12 E; 0 Z; 11 G; 3 H;  
SQ 8 I; 33 L; 8 K; 0 M; 4 F; 12 P; 14 S; 10 T; 4 W; 4 Y; 12 V;  
Initial Score = 20 Optimized Score = 44 Significance = 6.71  
Residue Identity = 26% Matches = 53 Mismatches = 124  
Gaps = 25 Conservative Substitutions = 0  
X 10 20 30 40 50 60  
MELTELLVWMLLTARLT--SSRAP--ACDLRVLSKLRLDSVHLSRLSQCEVHPLEPTVLLPAND  
IGVHCPCPAMLWLLSLVSLPLGLVPRGAPRLICDSVLERVYLLEAKQENITTTGCAEHCISINENITVDDTK  
X 10 20 30 40 50 60 70  
FSLGEKTMETKADILGAVTLLEGVMAARQGLPTCLSSLLGOLSGQVRLLIGALQSLTQLPQG-  
-LQSLIGTLPQGRKTAHKDPNALFTSFQHLRKGVRFLMVLVGGSTLCVRRAPPTTAVPSRSTVITLWELNRT  
EALISPDASAPLRTITADTFCKLFRVSNFLRGK---LKIYTG-EACRRGDR  
X 130 140 150 160 170 180 190  
VNFYAKRREVAQAQVEVWOGIALISEAV--LRGQAVLVNSSQPFEPQLQHMKAISGLRSTLTLRAIGAQ  
X 150 160 170 180 190  
5. US-08-223-263-1 (1-353)  
R28150 Sugar beet chitinase 1.  
ID R28150 standard; Protein; 439 AA.



VLCAIGMTALMAATTHIFRPLAKRCLOPDAARLPESAKLIFLYACMSYCAVILIG--TSYFFHDPDS  
70 80 90 100 110 120 130  
-----140 150 160 170 180 190  
-----OGRTAHKDPNAIFL--SF-QHLRGKRVFLMVGSTLCVRRAPPTTAVPSRTSVLTINELPNR  
VFYDMSGAAVWDJAVAYVILGSPYCSHAYATVMDSWRKDS-----VMLVHHVTL-----  
140 150 160 170 180  
200 210 220 230 240 250  
TSGLETFNTAARTTSGSL-----KWQGF-----RAKIPGLLNOTSSLDQIPGYINRIHELINGT  
190 200 210 220 230 240 250  
---LIASA-FFRYHNVGLVFLHDVSDVQLFETKINITYFKARGAGVHRLGLVANIGLSLFCFCWFWR  
260 270 280 290 300 310 320  
RGLFPGPSHRTIGADPDISGTS--DTGSLPNNLOPGYSPSETHPTGQYTLFPLPTLPTPVQHLPL--LP  
190 200 210 220 230 240 250  
LWVFP---LKVLATCCHSLQSVDPIDPYFFFFNILLIMVNNIYWFLYIVAFARVLTGOMHELEDLEXYD  
260 270 280 290 300 310 320  
330 340 X  
DPSAPPTP---TSPLLNTSYTHSQNSIOEG  
190 200 210 220 230 240 250  
TLEACTAKCKAEKPLRNGIVKDLF  
330 340 350  
7. US-08-223-263-1 (1-353)  
R45155 B.steatothermophilus DNA polymerase.

ID R45155 standard; Protein; 876 AA.  
AC R45155; 16-JUN-1994 (first entry)  
DE B.steatothermophilus DNA polymerase.  
KW Bacillus steatothermophilus; DNA polymerase; thermal stability;  
heat resistant; plasmid pDIF101.  
OS Bacillus steatothermophilus (IAM1001).  
PN J05304964-A.  
PD 19-NOV-1993.  
PF 27-APR-1992; 131400.  
PR 27-APR-1992; JP-131400.  
PA (TAKI ) TAKARA SHUZO CO LTD.  
DR WPI: 93-408323/51.  
N-PEDB: Q54170.  
PT DNA polymerase gene - is isolated from plasmid pDIF101, useful as  
reagent for gene engineering research  
PS Example 2; Page 8-11; 11pp; Japanese.  
CC DNA was extracted from Bacillus steatothermophilus IAM1001 and  
subjected to PCR amplification using the primers Q54168 and Q54169.  
CC A heat-resistant polymerase (R45155) can be recombinantly produced  
by transforming E.coli host cells with the gene comprising sequence  
CC Q54170.  
SQ Sequence 876 AA; 49 D; 0 B; 2 C; 38 Q; 90 E; 0 Z; 44 G; 17 H;  
83 A; 55 R; 25 N; 22 M; 33 F; 37 P; 37 S; 40 T; 5 W; 28 Y; 62 V;  
48 I; 99 L; 62 K; 22 M; 33 F; 37 P; 37 S; 40 T; 5 W; 28 Y; 62 V;  
Initial Score = 18 Optimized Score = 60 Significance = 5.87  
Residue Identity = 21% Matches = 84 Mismatches = 243  
Gaps = 69 Conservative Substitutions = 0

X  
MELTEL-LVVMLLT-ARLTLSPAPPADLAVLSKLRLDSVHLSRL-----SQCEV--  
GEEKVISDRLQGLASPHVVDITKRGITDIEPPYP-EYVERKYGLPPEQIVDIKGMGDSNIPVVG  
130 X 140 150 160 170 180 190  
60  
HPLPTEVLL-----PAVDFSLGEWK-TQMEETKADILGATVLLLEGMAARGDLPCTLSSILGDSQ  
190 200 210 220 230 240 250  
IEKTRAVKLLRQFTGVENVLASIDEIKGEKIKETLRQHRMA---LISKKLAIRDDAVELSLDDIAYGE  
200 210 220 230 240 250  
VRLLGATLSLIGTQLPPQRTTAHKDPNAIFLSFQHLRGKRVFLMIVGSLCY-----RRAP-PTTA  
120 130 140 150 160 170 180  
DREKVAALFKELGFSQSLERKESSESEKPLAKMAFTLADRVTEEMLDKALVVEVENEYHDAPVIGIA  
270 280 290 300 310 320 330  
V---PSRTSL-VLTINELPNRTSG-LFT---NTASARTTSGSLIKWQGFRAKIPGLINTSSLDQIRG  
190 200 210 220 230 240 250  
VNEHERFLREPTALADPOFVAMLDGETKKKSFEDSKR-AAVALRW-----KGIEL---CGVSFD-----  
340 350 360 370 380 390  
YINRIHELINGTRGLFPGPSHRTIGADPDISGTSIDTQSLPNNLOPGYSPSETHPTGQYTL--FPLPPTLP  
190 200 210 220 230 240 250  
-----LILAAYLLDPAGVDVAAAARKKQYEAHVSDAVYGGKGAIRAVPDEPLAHLVRKAAAIWALE  
400 410 420 430 440 450  
320 330 340 350 X  
TP-VVQHLPLPD-PSAETPTTSPBL-----NTSYTHSQNSIOEG  
460 470 480 490 500 X 510  
RPLDELIRNEQDRLVLEQPLSSILAEMEFAGVKVDYTRLEOMGELEAQLRTV  
460 470 480 490 500 510  
8. US-08-223-263-1 (1-353)  
R28348 Bacillus caldotenax DNA polymerase I.  
ID R28348 standard; Protein; 877 AA.  
AC R28348;  
DE 07-APR-1993 (first entry)  
KW Bacillus caldotenax DNA polymerase I.  
KW PolI; polymerase chain reaction; strain YT-G; DSM406;  
plasmid pDIF101.  
OS Bacillus caldotenax.  
FH Key Location/Qualifiers  
FT Misc difference 1  
FT /note="corresponds to TTC codon"  
PN EP-517418-A.  
PD 09-DEC-1992.  
PF 27-MAY-1992; 304763.  
PR 03-JUN-1991; JP-157368.  
PR 27-NOV-1991; JP-318685.  
PR 24-FEB-1992; JP-072090.  
PR 23-FEB-1992; JP-073161.  
PR 06-APR-1992; JP-112400.  
PA (TAKI ) TAKARA SHUZO CO LTD.



PI Fujita K, Ishino Y, Kato I, Uemori T;  
DR WPI; 92-408872/50.  
DR N-PSDB; Q31650.  
PT Gene for PolI type DNA polymerase and cloning method - for the  
PT production of DNA polymerase in high yield.  
PS Claim 2; Page 13-15; 30pp; English.  
CC B. caldtenax YT-G was cultured and DNA was isolated. The DNA was  
CC amplified by PCR using primers based on conserved amino acid motifs  
CC found in the PolI sequences of E. coli, Phage T7, Thermus aquaticus  
CC and Streptococcus pneumoniae (see Q31648 and Q31649). The 600bp  
CC amplified fragment was used as a probe to screen Southern blots  
CC of B. caldtenax genomic DNA. Restriction mapping of three positive  
CC fragments (i.e. a 1.45kb HincII fragment, a 2.4kb HindIII fragment  
CC and a 2.1kb XhoI fragment) showed that the fragments overlapped.  
CC The three fragments (with unneeded portions eliminated) were  
CC ligated with the vector pTV118N to produce plasmid pU1101  
CC containing a ca. 3.5kb combined fragment. E. coli HB101 transformed  
CC with the recombinant plasmid pU1101 expressed a heat-resistant DNA  
CC polymerase deduced to have the amino acid sequence given here.  
CC See also R28340-R28347 and Q31651.  
SQ Sequence 877 AA;  
SQ 84 A; 54 R; 25 N; 48 D; 0 B; 2 C; 39 Q; 90 E; 0 Z; 44 G; 18 H;  
SQ 48 I; 99 L; 62 K; 22 M; 33 F; 39 P; 36 S; 39 T; 5 W; 28 V; 62 V;  
Initial Score = 18 Optimized Score = 60 Significance = 5.87  
Residue Identity = 21% Matches = 86 Mismatches = 235  
Gaps = 81 Conservative Substitutions = 0  
X  
MELTEL-LIVMALLIT-ARLTSSPAPACDILRVLSKILRDSVHLISRL-----SCPEV--  
GREVKYISDRLTOLASHVTVDTTKGIDIDIEYTP-EAVREKYGITEQIVDLKGLMGKSDNIPVPS  
130 X 140 150 160 170 180 190  
HPLEPPVLL-----PAVDFSLGEMK-TOMEETKAODILGAVTLLIEGVMAARQGLPTCLSSILQSLGQ  
60 70 80 90 100 110  
VRLILGALOSLIGTQLPPGCRTHAKDNALFLSFOLHKGKVFIMLVGGSTLCV-----RRAP-PTTA  
120 130 140 150 160 170 180  
DEKXVALKKEIGFOSFLEKESPSSEEEKFLAKMAFTLADRVTEEMADKALVEAVEENYHADPIVIGIA  
200 210 220 230 240 250 260  
V---PSRTSL-VLTINELNRTSGL-LET---NFTASARTTSGILKMOOGFRAKIPGLINOTSRSIDQIPG  
190 200 210 220 230 240  
VNEHERFLRETLADPQVAVMIGDETRKMSFDSKR-AAVALKW-----KIGEL-CGVTFD-----  
340 350 360 370 380 390  
YINRIHEILNRCGLFPGPSRTTICA-----PDISSGTSIDGSLPPNIOGYSPPSPHPPTGQYTL  
250 260 270 280 290 300  
-----LLIAYILDPAGVDDVAALAAKKOYEAVERD-----EAVYGGAKAKRAVDEPVLAEHIVKRAAI  
400 410 420 430 440 450  
FPI-PTTL--PPVVOIHPL--LPDPSAPPTPTSPILNTSYTHSONLSOEG  
310 320 330 340 350 X

9. US-08-223-263-1 (1-353)  
R14854 Protein associated with biochemical pathway involv  
ID R14854 standard; Protein; 900 AA.  
AC R14854;  
DT 30-JAN-1992 (first entry)  
DE Protein associated with biochemical pathway involving CAMP.  
KW RAS; oncogene; cancer; cyclic nucleotide phosphodiesterase.  
OS Homo sapiens.  
PN MO9116457-A.  
PD 31-OCT-1991.  
PF 19-APR-1991; U02714.  
PR 20-APR-1990; US-511715.  
PA (COLD-) COLD SPRING HARBOR.  
PI Wiegler MH, Colicelli J; J;  
DR WPI; 91-339841/46.  
DR N-PSDB; Q14643.  
PT Complementary screening for genes and prods. - e.g. RAS protein  
PT and CAMP, that modify, complement or suppress genetic defect and  
PT correct associated phenotypic alteration  
PS Disclosure; Page 145; 169pp; English.  
CC In the specification this sequence is given the SEQ ID NO. 2 and  
CC is additional to SEQ ID NO. 6 1-53, i.e. the specification contains  
CC two sequences denoted as SEQ ID NO. 2. The only reference to SEQ ID  
CC NO. 2 in the text is to a 10 base pair linker. The origin and  
CC identity of R14854 is therefore obscure. Other sequences  
CC in the specification were isolated from human glioblastoma cells  
CC and encode cyclic nucleotide PDs and RAS-related polypeptides.  
CC (They were isolated by their ability to complement or suppress  
CC genetic defects in a RAS protein).  
SQ Sequence 900 AA;  
SQ 80 A; 63 R; 27 N; 36 D; 0 B; 12 C; 47 Q; 68 E; 0 Z; 37 G; 27 H;  
SQ 28 I; 107L; 25 K; 26 M; 26 F; 83 P; 85 S; 45 T; 13 W; 16 Y; 49 V;  
Initial Score = 18 Optimized Score = 64 Significance = 5.87  
Residue Identity = 20% Matches = 82 Mismatches = 239  
Gaps = 74 Conservative Substitutions = 0  
X  
MELTEL-LIVMALLIT-ARLTSSPAPACD-LRVLSKILRDS-----HVLHSRLS-QCEV  
LINTNSLAIMNDSVLENHHLAVGFKL---QEDNCDFONISKROSLRKAVIDWVATDMSKMTLL  
470 480 490 500 510 520  
HPLETPVLLPAVDFSLGEMKTOEETKAODILGAVTLLIEGVMAARQGLPTCLSSILQ-----IS  
60 70 80 90 100 110  
ADLKTWETKRV-TSSGVLLLDNYSDRIO-----VLRMNVHCA-DLSPTPTLTVQMTDRJMAEFQO  
530 540 550 560 570 580 590  
GOVRLILGALOSLIGTQLPPGCRTHA--HKDPNALFLSFOLHKGKVFIMLVGGSTLCVRRAP--TTAVP  
120 130 140 150 160 170 180  
GDRERERGEIIS-----PMCDKHTASVEKSGVGFIDYIVHPIWETWADLVHPSAISROSPPPEESRGP  
GOVRLILGALOSLIGTQLPPGCRTHA--HKDPNALFLSFOLHKGKVFIMLVGGSTLCVRRAP--TTAVP  
120 130 140 150 160 170 180







CC Rabbit high affinity IL-8 receptor gene was isolated from rabbit  
CC peritoneal neutrophils and used as a source of poly(A)+ RNA, to  
CC produce a rabbit neutrophil cDNA library. 250,000 recombinant  
CC plagues were screened for those which hybridized to an antisense  
CC oligonucleotide (Q30015). This probe was designed based on the  
CC sequence derived from the second transmembrane domain of G-protein-  
CC coupled receptors. After tertiary screening, six plaques were  
CC isolated. The insert of one of these plaques, termed F3R was of 2.5  
CC kb in size. This insert was sequenced. The protein deduced from  
CC the F3R clone demonstrates that it belongs to the family of  
CC G-protein-coupled receptors. The deduced protein sequence  
CC indicates seven putative transmembrane segments. A human  
CC peripheral blood leukocyte lambda gtl1 cDNA library (5' stretch) was  
CC screened with a 652 bp EcoRI/BamHI fragment (including nucleotides  
CC -27 to 623) of the rabbit F3R clone. After tertiary screening  
CC several human clones which hybridized to the rabbit IL-8 probe  
CC were isolated. The insert of one such clone, termed 4AB was  
CC sequenced (Q30012).

CC Sequence 360 AA;  
SQ 22 A; 18 R; 14 N; 14 D; 0 B; 9 C; 6 Q; 12 E; 0 Z; 17 G; 8 H;  
SQ 25 I; 65 L; 13 K; 9 M; 22 F; 13 P; 30 S; 18 T; 6 W; 14 Y; 25 V;  
Initial Score = 17 Optimized Score = 39 Significance = 5.45  
Residue Identity = 23% Matches = 55 Mismatches = 125  
Gaps = 51 Conservative Substitutions = 0

X  
MELT-----ELLVMMLLTARLTSSPPAPAC--DLNVLSKLRDSDHVLH  
150 X 160 170 180 190 200 210  
DRLVIAVHATRLTQKRYIVKFTICLSWGLSLILLAPVLLFRVYSSNVSPACYEDMGNNTANWR---ML  
50 60 70 80 90 100 110  
SRISQCEVHPLEPTVLLPAVDFSL-GEWKTOM-EETKADILGAVTL--LEGVMAARGOLGPTCLSSILG  
120 130 140 150 160 170 180  
QLSGQVR---LILGALOSL--LGTQLPPOGRTTAHKDPNAITLSFOHLIRKGVRLIMVGSTLCVRAPP  
190 200 210 220 230 240 250 260 270 280  
RILPQ-SFGFIVPILLIMFCYGFILRLIFKAMGOKHARMVIFAVALIFILCMPEYNILADTLMRTQVI  
290 300 310 320 330 340  
OETCERHMHIDRALDATEILGILHSCNP-----LIVAFIGQKFRH---GLIKITAIHG---LISKDSIP  
180 190 200 X 210  
TTAVPSSRTSLVLTINELPRTISGLLETNFTASART  
---KDSRPSFV---GSSSGHTSTVL  
350 360

14. US-08-223-263-1 (1-353)  
R14838 Protein deduced from human Glioblastoma cell cDNA.

ID R14838 standard; Protein; 405 AA.  
AC R14838;  
DT 30-JAN-1992 (first entry)  
DE Protein deduced from human Glioblastoma cell cDNA.  
KW RAS; oncogene; cancer; cAMP.  
OS Homo sapiens.

PN W09116457-A.  
PD 31-OCT-1991.  
PF 19-APR-1991; D02714.  
PR 20-APR-1990; US-511715.  
PA (COLD-) COLD SPRING HARBOR.  
PI Wiegler MH, Colicelli JJ;  
DR WPI; 91-339841/46.  
DR N-PSDB; 014626.  
PT Complementary screening for genes and prods. - e.g. RAS protein  
PT and cAMP, that modify, complement or suppress genetic defect and  
PT correct associated phenotypic alteration  
PS Example 1B; Page 70; 169pp; English.  
CC In the specification the sequence from which this amino acid  
CC sequence was deduced is given the SEQ ID NO. 13 and  
CC is described as the cDNA insert of plasmid pJC99. The insert  
CC sequence shows no significant homology to previously isolated  
CC genes. Plasmid pJC99 in E.coli (ATCC 68599) is described as  
CC containing a human glioblastoma cell cDNA insert encoding a  
CC RAS-related polypeptide. N.B. many of the references in the text of  
CC the specification to particular sequences are incorrect, therefore  
CC caution is advised in accepting that this sequence is indeed pJC99 !

SQ Sequence 405 AA;  
SQ 39 A; 36 R; 2 N; 9 D; 0 B; 9 C; 25 Q; 25 E; 0 Z; 32 G; 10 H;  
SQ 4 I; 74 L; 10 K; 6 M; 7 F; 33 P; 40 S; 18 T; 2 W; 8 Y; 16 V;  
Initial Score = 17 Optimized Score = 71 Significance = 5.45  
Residue Identity = 25% Matches = 90 Mismatches = 196  
Gaps = 67 Conservative Substitutions = 0

X  
MELT-----ELLVMMLLTARLTSSPPAPACDLRVLKLRDSDHVLHSLSCPEV  
70 80 90 100 110 120 130  
RHRAGQVVDILTYRQGRPFLEGRQALSRARMLT-ALDPKIVSKR--RLHVLKSL-RCSTVL  
60 70 80 90 100 110 120  
HPLTPVVL-----LPVDFSLGEWKTOMETKADILGAVTL-----LEGVMAARGOLGPTCLSSILG  
140 150 160 170 180 190  
KPL-RTIILARRRLADGSLGLRAGCGLARARQCGAGFSLSPVELEFQVRQKLLQVLT-----Y  
120 130 140 150 160 170 180  
QLSGQVRILGCA---LQSLGTO-----LPPOGRTTAHKD-PNMTFLSFQHLIRKGVRLIMVGST  
190 200 210 220 230 240 250 260 270  
SPSAQVKKRLQCKLIMALTQEGEGSGADGFLPILSLILAHCDLPELILFETMSLELPSLITGGGY  
280 290 300 310 320 330  
LCVRAPPTTAVPSSRTSLVLTINELP---NRTSGLETNTTASARTGSGSLKMQGFRPAKIPGLINTGSS  
340 350 360 370 380 390 400  
LTSLSH---SLALISGLQAHTLPLSPVOELRRLSLWEQRRLPATYHCFQHL-----RVAVDPSGGCTSKT  
240 250 260 270 280 290 300 310 320 330 340 350 360 370 380 390 400  
LQDIPCYINRIHELNGT--RCLEPGPSRRITGAPDISGTSGLAPLQD--GVSPEPTLPPIGOYTLP  
340 350 360 370 380 390 400  
LAVPDASATINOCATKFRVQTQNTFRLLF-----YKQGYHRLPPEPMTGCPPIATSTTAGQGLRP  
310 320



[illegible]

## PARAMETERS

Similarity matrix	Unary	K-tuple
Mismatch penalty	1	Joining penalty
Gap penalty	1.00	Window size
Gap size penalty	0.05	
Cutoff score	2	
Randomization group		
Initial scores to save	45	Alignments to save
Optimized scores to save	0	Display context
		15
		10
		32

## SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	6	7	2.60
Times:	CPU		
	00:02:53.97		Total Elapsed 00:02:34.00
Number of residues:		2246834	
Number of sequences searched:		75511	
Number of scores above cutoff:		4091	

Cut-off raised to 5.  
Cut-off raised to 6.  
Cut-off raised to 7.  
Cut-off raised to 8.  
Cut-off raised to 9.  
Cut-off raised to 10.  
Cut-off raised to 11.  
Cut-off raised to 12.

The scores below are sorted by initial score. Significance is calculated based on initial score

A 100% identical sequence to the query sequence was found:

Sequence Name	Description	Length	Init. Opt. Score	Sig. Frame
1. S4S331	c-MET ligand - human	353	353	133.36 0

The list of other best scores is:

Sequence Name	Description	Init.	Opt.	Length	Score	Score	Sig.	Frame
---------------	-------------	-------	------	--------	-------	-------	------	-------

2.	S45330	thrombopoietin - mouse	336	194	252	72.25	0
3.	A37222	**** 7 standard deviations above mean **** mucin, tracheal (AMN-22) - hu	294	26	55	7.69	0
4.	A29746	**** 6 standard deviations above mean **** epidermal growth factor-bind	259	23	46	6.53	0
5.	P00452	extensin-like protein - Persi	122	22	30	6.15	0
6.	S45222	ferrichrome-iron transport pr	659	22	59	6.15	0
7.	S07318	ferrichrome-iron transport pr	659	22	59	6.15	0
8.	QYBP	pyruvate carboxylase (EC 6.4.	1178	22	68	6.15	0
9.	S46094	pyruvate carboxylase (EC 6.4.	1180	22	68	6.15	0
10.	S24407	formin isoform IV - mouse	1206	22	67	6.15	0
11.	S11515	formin - mouse	1468	22	67	6.15	0
12.	S28148	**** 5 standard deviations above mean **** erythropoietin - rat	192	21	44	5.77	0
13.	FOGMND	gag polyprotein - Moloney mur	468	21	35	5.77	0
14.	S22701	mannosyltransferase (EC 2.4.1	480	21	36	5.77	0
15.	FOWM	gag polyprotein - Moloney mur	538	21	36	5.77	0
16.	A48954	beta-mannanase - Caldocellum	1331	21	62	5.77	0
17.	U01686	extensin-like protein precurs	151	20	34	5.38	0
18.	S45211	8-dihydro-6-hydroxymethylpt	159	20	33	5.38	0
19.	A43325	7,8-dihydro-6-hydroxymethylpt	159	20	33	5.38	0
20.	A24902	erythropoietin precursor - mo	192	20	46	5.38	0
21.	U00173	erythropoietin precursor - cr	192	20	43	5.38	0
22.	A24901	erythropoietin precursor - mo	193	20	43	5.38	0
23.	S06686	protein rpx-VT3 - Thermoprote	268	20	31	5.38	0
24.	S13383	hydroxyproline-rich glycoprot	283	20	46	5.38	0
25.	S28264	hydroxyproline-rich glycoprot	323	20	44	5.38	0
26.	S11674	acrosin (EC 3.4.21.10) precu	401	20	68	5.38	0
27.	S45025	chitinase - beet	439	20	40	5.38	0
28.	S15921	protein rpx-VT3 - Thermoprote	474	20	61	5.38	0
29.	S07096	hexose transport protein - Ch	533	20	60	5.38	0
30.	FOWMV	gag polyprotein - Moloney mur	538	20	35	5.38	0
31.	S22700	amphipysin - chicken	682	20	59	5.38	0
32.	GNWTW	genome polyprotein - hepatitis	3010	20	68	5.38	0
33.	S26198	**** 4 standard deviations above mean **** H+-transporting ATP synthase	248	19	36	5.00	0
34.	S32896	hypothetical protein 9 - Salm	299	19	40	5.00	0
35.	B39363	GDF-1 embryonic growth factor	350	19	63	5.00	0
36.	S27200	proline-rich protein - mouse	389	19	26	5.00	0
37.	HXXR22	sigma 1 protein - reovirus ty	399	19	63	5.00	0
38.	C34829	sigma 1 protein - reovirus ty	462	19	63	5.00	0
39.	P2MTB	I2 protein - bovine papilloma	469	19	67	5.00	0
40.	S47076	beta-ketoacyl-ACP synthase -	489	19	61	5.00	0
41.	A33526	collagen alpha 2(IIV) chain pr	1707	19	73	5.00	0
42.	S17941	xpsi protein - Xanthomonas ca	1378	18	36	4.61	0
43.	A46132	c-Jun leucine zipper interact	149	18	30	4.61	0
44.	S37679	hypothetical protein 18 (coxI	165	18	38	4.61	0
45.	P00475	piscil extensin-like protein	171	18	30	4.61	0

```

1. US-08-223-263-1 (1-353)
   S45331
      C-MPI ligand - human
ENTRY
TITLE      S45331 #type complete
ORGANISM   C-MPI ligand - human
           #formal name Homo sapiens #common name man

```











Residue Identity = 24% Matches = 79 Mismatches = 170  
Gaps = 77 Conservative Substitutions = 0

```

X
MELTEL-----LVMLLTARLTSSPAPACDLRLSKLSDS-HVL-HSRLSQCEVHP
RDAGHWASGALLIEDLMPWRMPRLMALFAGVNL--AVAGCIIORLTGNPASEVELGISGAFAVVM
370 380 390 400 410 420 430
60
LPTPVLLPAVD-----SLGEKWTQMETKADIIAGAVTLLEGVMAARGDLPCTCISLLGQLSGVRL
L-----FLVPEGNAGWMLPAGSLG-----AAVTLIIIMLAGRGCFSPHMLLAGMALSTFTM
440 450
130 140 150 160 170
LIGALQSLIGTQLPPQ-----GRTTAHKD-----PNAIFLSFQHLRGKVRFLMVGSTLCVRRAP
LIMMLD-ASGDRMAQVLTWISGSTYNAITDAQWRGTGIWVTLTAITPLCRBMLTLPL--GGDT--AAVAG
500 510 520 530 540 550
180 190 200 210 220 230 240 250
TTAVPSRTSLVLTINELPNRTSGLETNFTASARTTSGILKMOGRAPRIPGILNQTSSISDQIPGYLMRI
MALTFRIAL-----LILAACILATATMTIGPL--SFVGLMAPHIAAMGFRRTMPEHI-----VI
560 570 580 590 600 610
260 270 280 290 300
HEIINSTRGLF-PGPSRRITGAPDISGSDTGSILP---NLQPGYSPSPTHPTGQY
SALVGLIIVFADWCGRMVLFEPQIPAGILSTFIQAPFYIILKQSR
620 630 640 650
X
8. US-08-223-263-1 (1-353)
QYBP pyruvate carboxylase (EC 6.4.1.1) 1 - yeast (Sacch
ENTRY QYBP #type complete
TITLE pyruvate carboxylase (EC 6.4.1.1) 1 - yeast (Saccharomyces
ALTERNATE NAMES pyruvic carboxylase
ORGANISM #formal name Saccharomyces cerevisiae
DATE 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
ACCESSIONS A29233; S05760; A29722
REFERENCE A92662
#authors Lim, F.; Morris, C.P.; Occhiodoro, F.; Wallace, J.C.
#journal J. Biol. Chem. (1988) 263:11493-11497
#title Sequence and domain structure of yeast pyruvate carboxylase.
#cross-references MUID:88298805
#accession A29233
#molecule_type DNA
#residues 1-1178 #label LIM
#cross-references EMBL:J03889
S05760
#authors Morris, C.P.; Lim, F.; Wallace, J.C.
#journal Biochem. Biophys. Res. Commun. (1987) 145:390-396
#title Yeast pyruvate carboxylase: gene isolation.
#cross-references MUID:87241529
#accession S05760
```

##molecule\_type DNA  
##residues 1003-1178 ##label MOR  
##cross-references EMBL:J03889  
#accession A29722  
##molecule\_type protein  
##residues 1124-1149 ##label MOR2

## GENETICS

#gene LISTA:PYC1

## CLASSIFICATION

#superfamily pyruvate carboxylase  
#biotin; gluconeogenesis; homotetramer; ligase; zinc

## KEYWORDS

FEATRE

157-331

353-468

569-908

1089-1178

SUMMARY

SEQUENCE

#domain ATP/bicarbonate-binding #label ATB1  
#domain ATP/bicarbonate-binding #label ATB2  
#domain pyruvate-binding #label PYR  
#domain biotinyl- or lipoyl-binding #label BIO  
#binding site biotin (lys) (covalent) #status predicted  
#length 1178 #molecular-weight 130098 #checksum 6631

Initial Score = 22 Optimized Score = 68 Significance = 6.15  
Residue Identity = 23% Matches = 92 Mismatches = 229  
Gaps = 79 Conservative Substitutions = 0

```

X
MELTELLVWMLLTARLTSSPAPACDLRLVLS-----KLSDSHVLSRLSQCEVHPILP
RINQVKTNIFLLITLNPVFIEGTWGT-FIDDTPLQFQWSSQNRARQLHYLADVADNGSSIKQIGLP
440 X 450 460 470 480 490 500
60 70 80 90 100 110 120
TPVILPAVDPSLGEKWTQMETKADIIAGAVTLLEGVMA--ARGQLPCTCISLLGQLSGOVR-----LI-
KIKSNPVSPEHLHDAQGNVINVTKSAPESGMQVILKRGPAEFAR-----QVRQPNGTLLA
510 520 530 540 550 560
130 140 150 160 170 180
-----LIGALQSLIGT-----QLPEQRTTAHKDPNAIFLSFQHLRGKVR--RFLMVGSGSTLCVRRAPPTTV
DTTWRAHOSLATAIKRTHDIAITAPITTAHLAGPALECGAGTIPVYANRFLHEDPWQRLKLR-----SLV
570 580 590 600 610 620
190 200 210 220 230
PSRTSLVLTINELPNRTSGL--LETNFTASA-----RTTSGILKMO--QGFR--KIPGILNOT-SR
PNIIPFQMLRGANGVAYSSLPDPAIDHFVKQAKONSVDIFRPDALNDLQIAKVGVDAYKAGGVVEATVCF
630 640 650 660 670 680 690 700
240 250 260 270 280 290 300
SLD-QIPC-YLMRIHEILNCTRGLFPGPSRRITGAPDISGCT--SPTGSLPNLQPGSPSPTHPTGQYT
SGMDLPQKRYNDLYLEIAEKIVQMG--THILGIRD-MAGTKPAPAAKLILGSLIRAKIPDLP IHVT-----
710 720 730 740 750 760
310 320 330 340 350 X
LFLPPTIIPVQHLPLPDPSAFTPTPTSPILNTSYTHSONL--SQEG
-HDSAGIRVASMTACALACADVADVAINSMGSL--TSQPSINALLASLEGIDTGINVEH
770 780 790 800 810 X 820
```

9. US-08-223-263-1 (1-353)  
S46094 pyruvate carboxylase (EC 6.4.1.1) 2 - yeast (Sacch

ENTRY #type complete  
TITLE pyruvate carboxylase (EC 6.4.1.1) 2 - yeast (Saccharomyces cerevisiae)  
ALTERNATE\_NAMES protein YBR1507; protein YBR218C  
ORGANISM #formal name Saccharomyces cerevisiae  
DATE 26-Aug-1994 #sequence\_revision 09-Sep-1994 #text\_change 14-Sep-1994  
ACCESSIONS S46094; S17469  
REFERENCE S45782  
#authors Dubois, E.; El Bakoury, M.; Glansdorff, N.; Messenguy, F.; Pierard, A.; Scherens, B.; Vierendeels, F.  
#submission submitted to the Protein Sequence Database, August 1994  
#accession S46094  
#molecule\_type DNA  
#residues\_type DNA 1-1180 #label DUB  
#cross-references EMBL:236087  
REFERENCE S17469  
#authors Stucka, R.; Deguin, S.; Salmon, J.M.; Gancedo, C.  
#journal Mol. Gen. Genet. (1991) 229:307-315  
#title DNA sequences in chromosomes II and VII code for pyruvate carboxylase isoenzymes in Saccharomyces cerevisiae: analysis of pyruvate carboxylase-deficient strains.  
#cross-references MUID:9201767  
#accession S17469  
#molecule\_type DNA  
#residues 1-14 'C', 16-131, 'E', 133-237, 'K', 239-267, 'F', 269-545, 'C', 547-641, 'T', 643-770, 'S', 772, 'R', 774-830, 'R', 832-838, 'P', 840-1000, 'N', 1002-1134, 'R', 1156-1177, 'P', 1179-1180, 'V', 'I', 'F', 'R' #label STU  
#cross-references EMBL:X59890

GENETICS  
#gene LISTA:PYC2  
#map position 2R  
CLASSIFICATION #superfamily pyruvate carboxylase  
KEYWORDS biotin; gluconeogenesis; homotetramer; ligase; zinc  
FEATURE  
1136  
#binding site biotin (lys) (covalent) #status predicted  
SUMMARY #length 1180 #molecular-weight 130166 #checksum 1031  
SEQUENCE

Initial Score = 22 Optimized Score = 68 Significance = 6.15  
Residue Identity = 23% Matches = 91 Mismatches = 237  
Gaps = 65 Conservative Substitutions = 0

X 10 20 30 40 50  
MELTELLVMLLTARLTSSPAPACDILRVLS-----KLIDSHVLSRLSQCEVHP  
RIRGVKTNIPFLITLTNPVFIEGT-ywTTFIDTFQOLFQWSSONRAOKLIHVIADLAVNGSSIKGQIGLP  
440 X 450 460 470 480 490 500  
TPPLPAVDVFSIGEKTKQETKADIIIGAVTLLIEGMAARGQLGPTCLSSLLGQLSGVRL-----LGA  
KKSNSVPHLHDAQGNVINTKSAFPGSGMROVLE-----KGPSEFAKQVROFNG--TILMDTWRDA  
510 520 530 540 550 560

10. US-08-223-263-1 (1-353)  
S24407 formin isoform IV - mouse

ENTRY #type complete  
TITLE formin isoform IV - mouse  
ORGANISM #formal name Mus musculus #common name house mouse  
DATE 19-Feb-1994; #sequence\_revision 19-Feb-1994; #text\_change 19-Feb-1994  
ACCESSIONS S24407  
REFERENCE S24407  
#authors Jackson-Grusby, L.; Kuo, A.; Leder, P.  
#journal Genes Dev. (1992) 6:29-37  
#title A variant limb deformity transcript expressed in the embryonic mouse limb defines a novel formin.  
#cross-references MUID:92112033  
#accession S24407  
#status preliminary  
#residues 1-1206 #label JAC  
#cross-references EMBL:X62379  
SUMMARY #length 1206 #molecular-weight 133463 #checksum 501  
SEQUENCE

Initial Score = 22 Optimized Score = 67 Significance = 6.15  
Residue Identity = 23% Matches = 95 Mismatches = 220  
Gaps = 95 Conservative Substitutions = 0

X 10 20 30 40 50  
MELTE-----LLIVVMLLTARLTSSPAPACDILRVLSKLIDSHVLSRLSQCE  
EQSLQNLIDMPRTQKRADEPFHGADMGVSTQESHKSP-RDAHVGGGVKARFETALAFALFIRP-  
350 X 360 370 380 390 400 410  
VHLPPTVLLPAVDPS-LGEKTKQETKADIIIGAVTLLIEGMAARGQLGPTCLSSLLGQLSGVRLILG  
60 70 80 90 100 110 120

-----PKKSTADTSELEALKRKKHEK--ESLRNV-----FEKSRPADSESDKSPDQSEODDRTPG  
420 430 440 450 460 470  
ALQ-----SLGTQLPPGRTTAHKNDAIFLSFOHLR-GKVRFLMVGSTLCVRRAPTTA-VPSRTS  
130 140 150 160 170 180  
RLOAWPPPKTDEKVGKYTEAEYQAI-----LHKREKEIEITLQAEFLKTHHNGEHALVTAR--  
480 490 500 510 520 530  
190  
LVTITNE-----LPRNTSGILE-TNFTASARTGS--GLKMOGFR--AKIPGLNQTSSIDQIP  
200 210 220 230 240  
-----LEAELENKQOLEKREGEEMRDVCISTDDCSKPAFRNVCIOTRTEFLPCAEKATRS-SQI-  
540 550 560 570 580 590 600  
GYINRHEILNCTRGILPPEPSRRTLGAP-DISSGTS-----DTGSLP-----PNLQGYSPSPTHTPTGQYTL  
250 260 270 280 290 300  
-----VPRKLTITSLTQLSPSKSDIHAFQTRREGTSSSQKISPPAPPPPPPLPPLIPPPPLPPG---L  
610 620 630 640 650 660  
310 320 330 340 350 X  
FLPPLTPPVV--QLHPLLPDSAPTPPTS-----PLNFTSYTHSQNLISQEG  
670 680 690 700 710 720 X 730  
GPIPPAPPIPPVCPVSPPPPPPTTPVPSDGPDPPPPPPLPN-----VIALPNSGGGPPPPPPPP  
670 680 690 700 710 720 X 730

11. US-08-223-263-1 (1-353)  
S11515 formin - mouse

ENTRY S11515 #type complete  
TITLE formin - mouse  
ORGANISM #formal name Mus musculus #common name house mouse  
DATE 22-Jan-1994; #sequence\_revision 22-Jan-1994; #text\_change 22-Jan-1994  
ACCESSIONS S11515  
REFERENCE S11515  
#authors Woychik, R.P.; Maag, R.L.; Zeller, R.; Vogt, T.F.; Leder, P.  
#journal Nature (1990) 346:850-853  
#title 'Formins': proteins deduced from the alternative transcripts of the limb deformity gene.  
#accession S11515 preliminary  
#status 1-1468 #label WOY  
#residues 1-1468 #label WOY  
#cross-references EMBL:X53599  
SUMMARY #length 1468 #molecular-weight 163808 #checksum 6099  
SEQUENCE

Initial Score = 22 Optimized Score = 67 Significance = 6.15  
Residue Identity = 23% Matches = 95 Mismatches = 224  
Gaps = 86 Conservative Substitutions = 0  
X 10 20 30 40 50  
METTELLVWMLLTARLT-LSPAPACDRLVLSKLLRDSHYL-----HSRLSOCEVHEPLP  
SADTLEPSSITVTETKASPTSLASQTWVSEASKEIGEPKRTAPRQHLPEGLASGPPC-DNKKEQ  
570 580 590 600 610 620 630  
60 70 80 90 100 110 120

TPVILPAVDLSIGEWKQMEETKAQDILIGAVTLLIEVMNARGQLGPTCISLLGO-LSGQVRLILALQ--  
130 140 150 160 170 180 190  
TAKDLPKND--GGVWVPYGRAGPCCPL-----LHEEKKSRSSEL---YIDLNPDSQPTQODRRTPGLQAV  
640 650 660 670 680 690 700  
-----SLGTQLPPGRTTAHKNDAIFLSFOHLR-GKVRFLMVGSTLCVRRAPTTA-VPSRTSIVLT  
130 140 150 160 170 180 190  
WPPPKTDEKVGKYTEAEYQAI-----LHKREKEIEITLQAEFLKTHHNGEHALVTAR-----L  
710 720 730 740 750 760  
NE-----LPRNTSGILE-TNFTASARTGS--GLKMOGFR--AKIPGLNQTSSIDQIPGYLNR  
200 210 220 230 240  
-----LEAELENKQOLEKREGEEMRDVCISTDDCSKPAFRNVCIOTRTEFLPCAEKATRS-SQI-  
770 780 790 800 810 820 830  
EAELENKQOLEKREGEEMRDVCISTDDCSKPAFRNVCIOTRTEFLPCAEKATRS-SQI-  
770 780 790 800 810 820 830  
250 260 270 280 290 300 310  
IHEILNCTRGILPPEPSRRTLGAP-DISSGTS-----DTGSLP-----PNLQGYSPSPTHTPTGQYTL  
610 620 630 640 650 660 670  
PKKLTITSLTQLSPSKSDIHAFQTRREGTSSSQKISPPAPPPPPPLPPLIPPPPLPPG---LAPLTP  
840 850 860 870 880 890  
320 330 340 350 X  
TLETPVV--QLHPLLPDSAPTPPTS-----PLNFTSYTHSQNLISQEG  
900 910 920 930 940 950  
APPPIPPVCPVSPPPPPPTTPVPSDGPDPPPPPPLPN-----VIALPNSGGGPPPPPPPP  
900 910 920 930 940 950

12. US-08-223-263-1 (1-353)  
S28148 erythropoietin - rat

ENTRY S28148 #type complete  
TITLE erythropoietin - rat  
ORGANISM #formal name Rattus norvegicus #common name Norway rat  
DATE 22-Nov-1993; #sequence\_revision 22-Nov-1993; #text\_change 22-Nov-1993  
ACCESSIONS S28148  
REFERENCE S28148  
#authors Nagao, M.; Suga, H.; Okano, M.; Masuda, S.; Narita, H.;  
#journal Biochim. Biophys. Acta (1992) 1171:99-102  
#title Nucleotide sequence of rat erythropoietin.  
#accession S28148 preliminary  
#status 1-192 #label NAG  
#residues 1-192 #label NAG  
SUMMARY #length 192 #molecular-weight 21286 #checksum 4587  
SEQUENCE

Initial Score = 21 Optimized Score = 44 Significance = 5.77  
Residue Identity = 27% Matches = 53 Mismatches = 118  
Gaps = 24 Conservative Substitutions = 0  
X 10 20 30 40 50 60  
METTELLVWMLLTARLTSSPAP--ACDILRVLSKLLRDSHYLHSRLSOCEVHEPLP  
MGVPERITLLLSILLPLIGLVLCAPRLLICDSRVLEYIIIEAKEAEENVMTGCAEGRLSENITVDTTV  
X 10 20 30 40 50 60 70







```

160      170      180      190      200      210      220
RG-----KVFIMLVGSGTLCVBRAP-PTTAVPSRTSLVLTNLEPNRTSGILETNFTASARTGSGILK
|         |         |         |         |         |         |         |
SGSYSDYVYSLISDSDSIIFPLENIXAVPIVIOAFKLMFGNIIKADFLXLLIFARGGISDMOT--MLLK
130      140      150      160      170      180      190
230      240      250      260      270      280      290
WQGFRAKIPGLINOTSRSLDQIPGVLNRIHELL-NGTRGLFPGPSRRTLGADISSGTSLEPNIOPG
|         |         |         |         |         |         |         |
PIDSFSPQSKSMWLNIIIDLNKPIP-KYKSKPSLSSDEISHQPELVICADPRDDWSEWARRQFCQWT
200      210      220      230      240      250      260
300      310      320      330      340      350
YSPSPHPTGGYTLFPLEPTLPFVVOLEHPLP-----DP-----SAPFTPTSPILNTSYTHS---QNLISQ
|         |         |         |         |         |         |         |
IOAKPEHPIRLLEIILNITATTL--ASVO-NGGVVSESIDPREEDYNNVYRKRRADEITYKASELKNKNV
270      280      290      300      310      320      330
X
EG
|
DGSDDINNTGPG
X 340

15. US-08-223-263-1 (1-353)
FOMVM      gag polyprotein - Moloney murine sarcoma virus

ENTRY      FOMVM      #type complete
TITLE      gag polyprotein - Moloney murine sarcoma virus
CONTAINS   core protein p15; core shell protein p30; inner coat protein
           p12; nucleoprotein p10
ORGANISM   #formal name Moloney murine sarcoma virus
DATE       18-Dec-1981 #sequence_revision 18-Dec-1981 #text_change
           30-Sep-1993
ACCESSIONS A94261; B00645; A03932
REFERENCE   A94261
#authors    Reddy, E.P.; Smith, M.J.; Aaronson, S.A.
#journal    Science (1981) 214:445-450
#title      Complete nucleotide sequence and organization of the Moloney
           murine sarcoma virus genome.
#cross-references M01D:82039539
#contents   Provirus
#accession  A94261
#molecule_type genomic RNA
REFERENCE   A00645
#residues   1-538 ##label RED
#authors    Van Beveren, C.; van Straaten, F.; Gallethaw, J.A.; Verma,
           I.M.
#journal    Cell (1981) 27:97-108
#title      Nucleotide sequence of the genome of a murine sarcoma virus.
#cross-references M01D:82113347
#contents   Clone 124, circular
#accession  B00645
#molecule_type DNA
#residues   1-518, 'K', 520-538 ##label VAN

GENETICS
#gene       gag
CLASSIFICATION #superfamily mammalian retrovirus gag polyprotein I
FEATURE

```

```

2-131      #product core protein p15 #label P15\
132-215    #product inner coat protein p12 #label P12\
216-478    #product core shell protein p30 #label P30\
479-534    #product nucleoprotein p10 #label P10
SUMMARY    #length 538 #molecular-weight 61209 #checksum 590
SEQUENCE

Initial Score = 21 Optimized Score = 36 Significance = 5.77
Residue Identity = 24 Matches = 43 Mismatches = 114
Gaps = 21 Conservative Substitutions = 0

170      180      190      200      210      220      230
TLGVRAPPTTAVPSRTSLVLTNLEPNRTSGILETNFTASARTGSGILK---WQGFRAKIPGLINOTSR
|         |         |         |         |         |         |         |
MGQVTTPLSLTL--DHMKVERLAH-----QSDVYKRRRWTFCSALWP-----TTN
X         10         20         30         40

240      250      260      270      280      290      300      310
SLDQIPGVLNRIHELLNGTRGLFPGPSRRTLGADISSGTSLEPNIOPGYSPTHPPTGGYTLFPLP
|         |         |         |         |         |         |         |
VGPFRDGTFRNBDLITQVAKIKVTFSPGHCHDPQVPIYTWALAFDPPPWKPFVHKKPPPIILPSAPSLPE
50      60      70      80      90      100      110

320      330      340      350      X
PLIPLT-PVVOLEHPLDPSAPPTPTSPILNTSYTHSONLSQG
|         |         |         |         |         |         |         |
PLSLTPQSSLYPAL-TPS-IGAKPKQVLSDSGCPILIDLTEDPPRYRDRRP
120      130      140      150      160      170

> 0 <
0 1 10 Intelligenetics
> 0 <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file sqslpt.res made by on Tue 11 Apr 95 9:56:26-PDT.

Query sequence being compared:US-08-223-263-1 (1-353)
Number of sequences searched: 40292
Number of scores above cutoff: 4111

Results of the initial comparison of US-08-223-263-1 (1-353) with:
Data bank : Swiss-Prot 30, all entries

10000-      *      *
N      -      *      *
U 5000-      *      *
M      -      *      *
B      -
E      -
R      -      *      *
O      -      *      *
F 1000-      *      *
S      -      *      *
E 500-      *      *
Q      -

```







GN FHOB.  
OS ESCHERICHIA COLI.  
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;  
OC ENTEROBACTERIACEAE.  
RN (1)  
RP SEQUENCE FROM N.A.  
RM 87014116  
RA KOESTER W., BRAUN V.;  
RL MOL. GEN. GENET. 204:435-442(1986).  
RN (2)  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / W3110;  
RM 94261430  
RA FUJITA N., MORI H., YURA T., ISHIHAMA A.;  
RL NUCLEIC ACIDS RES. 22:1637-1639(1994).  
RN (3)  
RP SEQUENCE OF 1-23 FROM N.A.  
RM 88038363  
RA BURKHARDT R., BRAUN V.;  
RL MOL. GEN. GENET. 209:48-55(1987).  
CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS INVOLVED IN THE HIGH-  
CC AFFINITY TRANSPORT OF IRON(III)-FERRIC IRON INTO THE E.COLI  
CC CELL.  
CC -1- SUBCELLULAR LOCATION: INNER MEMBRANE-ASSOCIATED.  
CC -1- SIMILARITY: TO THE FECC AND FECD PROTEINS OF THE IRON (III)  
CC DIGITATE TRANSPORT SYSTEM.  
DR EMBL; X04319; ECFHDB.  
DR EMBL; D26562; ECFHDB.  
DR EMBL; X05810; ECFHDB.  
DR PIR; S07318; S07318.  
DR ECOCENE; EG10303; FHUB.  
KW IRON TRANSPORT; TRANSPORT; INNER MEMBRANE; SIGNAL.  
FT SIGNAL 1 20  
FT CHAIN 21 659 FERRIC IRON TRANSPORT PROTEIN FHOB.  
FT CONFLICT 11 11 L -> V (IN REF. 3).  
SQ SEQUENCE 659 AA; 70335 MW; 2233995 CN;

Initial Score = 22 Optimized Score = 59 Significance = 7.74  
Residue Identity = 24% Matches = 79 Mismatches = 170  
Gaps = 77 Conservative Substitutions = 0

X 10 20 30 40 50  
MELTELL-----LVMLLTARLTSPAPACDRLVLSKLRDS-HVL-HSRLSCEVHE  
ROAHGWTWASGLIEDLMPWRMPRIALALFAGVNL---AVAGCITORTLTPNPAASEVLGSSGAFGVLLML  
370 380 390 400 410 420 430  
60 70 80 90 100 110 120  
LPTPVLLPAVDF-----SLEEMKTOEETKRAODILGAVILLLEGMAARAGOLGFTCLSLTGLSGGVR  
L---FLVPGNARGMILPASISG-----AAVTLILMTAARGRGSPHRMLIAGMALSTFTT  
440 450 460 470 480 490

130 140 150 160 170  
ILGALGSLTGLTPQ-----GRTTAKRD-----PNAIFLSOHLRGVRLMVGSLTCVRRAPE  
11  
LIMMLD-ASGDDPRMAOVLTWIGSGTYNATDQAVWRGTGLVITLLAITPLCRMLTLLPLT-GGDT---AAVAG  
500 510 520 530 540 550

```

5. US-08-223-263-1 (1-353)
PYCL_YEAST PYRUVATE CARBOXYLASE 1 (EC 6.4.1.1) (PYRUVIC CARBO
ID PYCL YEAST STANDARD; PRT: 1178 AA.
AC P11154;
DT 01-JUL-1989 (REL. 11, CREATED)
DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)
DE 01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
DE PYRUVATE CARBOXYLASE 1 (EC 6.4.1.1) (PYRUVIC CARBOXYLASE 1) (PCB 1).
GN PYCL OR PYV.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RM 88298805
RA LIM F., MORRIS C.P., OCCHIODORO F., WALLACE J.C.;
RL J. BIOL. CHEM. 263:11493-11497(1988).
RN [2]
RP SEQUENCE OF 1003-1178 FROM N.A.
RM 87241529
RA MORRIS C.P., LIM F., WALLACE J.C.;
RL BIOCHEM. BIOPHYS. RES. COMMUN. 145:390-396(1987).
CC -1- FUNCTION: PYRUVATE CARBOXYLASE CATALYZES A 2-STEP REACTION,
INVOLVING THE ATP-DEPENDENT CARBOXYLATION OF THE COVALENTLY
ATTACHED BIOTIN IN THE FIRST STEP AND THE TRANSFER OF THE
CARBOXYL GROUP TO PYRUVATE IN THE SECOND.
CC -1- CATALYTIC ACTIVITY: ATP + PYRUVATE + HCO(3) (-) = ADP +
CARBOXYLPHOSPHATE + OXALACETATE.
CC -1- PATHWAY: GLUCONEOGENESIS.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- COFACTOR: BIOTIN, AND ZINC.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: WITH OTHER BIOTIN CARBOXYLASES, LIPOMIDE
TRANSFERASES AND CARBAMYL PHOSPHATE SYNTHETASES.
DR EMBL; J38889; SFCB.
DR PIR; A29233; QYBP.
DR PROSITE; PS00188; BIOTIN.
DR PROSITE; PS00866; CPSASE_1.
DR PROSITE; PS00867; CPSASE_2.
KW LIGASE; MULTIFUNCTIONAL ENZYME; BIOTIN; GLUCONEOGENESIS; ZINC;
KW MULTIGENE FAMILY.
FT 182 187 ATP (POTENTIAL).
FT NP BIND 312 312 BY SIMILARITY.
FT BINDING 1135 1135 BIOTIN (BY SIMILARITY).
FT SIMILAR 160 330 CARBAMOYL PHOSPHATE SYNTHETASES.
FT SIMILAR 350 470 WITH OTHER BIOTIN CARBOXYLASES.
FT SIMILAR 1086 1178 WITH OTHER BIOTIN CARRIER PROTEINS AND
WITH LIPOMIDE ACETYLTTRANSFERASES.

```

SQ SEQUENCE 1178 AA; 130099 MW; 7059028 CN;

Initial Score = 22 Optimized Score = 68 Significance = 7.74  
Residue Identity = 23% Matches = 92 Mismatches = 229  
Gaps = 79 Conservative Substitutions = 0

X 10 20 30 40 50  
MELTELLVYMLLTLARLTSSPAPACDLNVLS-----KILNDSVHLSRLSQCEVHPLE  
|||  
RIRGVKTINFLITLTLPNEVFIEGTWGI-FIDTPQLFQMWSQNPAOKLTHLYADVDNGSSIKQGLIF  
440 X 450 460 470 480 490 500  
60 70 80 90 100 110 120  
TPVLPLAVDFSLSECKTQWEERKADILLGATVLTLEGYMA--ARGQLGPCTLSSLIGQLSGVR-----LI-L-  
|||  
KLKSNDVPHLHDAGGNVINTKSAPEPSGMROVLKEKPAEFAR-----QVRPNGLTILM  
510 520 530 540 550 560  
130 140 150 160 170 180  
----LGALISLIGT-----QLPPGRRTAHDPNAIFLSPGHILRGKV--RFIMLVGGSTLCVRAPPTVAIV  
|||||  
DTTWKDASHOSLTAIRTRTDIATIPTAHALAAGFALECGGATFPVANRFLHEPDWRQLKLR-----SLVV  
570 580 590 600 610 620  
190 200 210 220 230  
PSRTSLVLTINLETNPRTSCL--LETNPNTAS-----RTTGSGILKM--QGFR-A-KIPGLLNQT-SRS  
|||  
PNIPFMOLLRGANGVAYSLLPONAIDHFVKQAQNSVDIFPFVALNDLEQLKVGDAVKKAGGYEATVCFE  
630 640 650 660 670 680 690 700  
240 250 260 270 280 290 300  
SLD-QIQG-YINRIEHLINGTSGLPFGPSRRILGAPDISGT---SDTGSIPNLIGQVSPTSPTPTPGQXT  
|||  
SGDMLOPKKNINDYLTLEAEKIVOMG--THLIGKD-MAGTMPPAAKKILLISIRAKTIPDLPIHVHT-----  
710 720 730 740 750 760  
310 320 330 340 350 X  
LFPLPPLTPPVVQLHPLLPDPSPAPTPTSPILTNTSYTHSONI--SQEG  
|||  
-HDSAGTRVASMTACHLAGADVVDVAISMSEL--TSQPSINALLASLEFNIDTGINVEH  
770 780 790 800 810 X 820  
6 US-08-223-263-1 (1-353)  
PYC2\_YEAST PYRUVATE CARBOXYLASE 2 (EC 6.4.1.1) (PYRUVIC CARB

ID PYC2\_YEAST STANDARD; PRI; 1180 AA.  
AC P32327;  
DT 01-OCT-1993 (REL. 27, CREATED)  
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)  
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)  
DE PRIVATE CARBOXYLASE 2 (EC 6.4.1.1) (PYRUVIC CARBOXYLASE 2) (PCB 2).  
GN PYC2 OR YBR218C OR YBR1507.  
OC SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMETES.  
RN  
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
RM 92017667  
RA STUCKA R., DEQUIN S., SAUMON J.-M., GANCEDO C.;  
RL MOL. GEN. GENET. 229:307-315(1991).

RN SEQUENCE FROM N.A.  
 RC STRAIN-5288C;  
 RA DOBOSI E., EL BAKOORY M., GLANDORFF N., MESSENGUY F., PIERARD A.,  
 BA SCHEERENS B., VIERENDELS F.;  
 RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
 CC -1- FUNCTION: PYRUVATE CARBOXYLASE CATALYZES A 2-STEP REACTION,  
 INVOLVING THE ATP-DEPENDENT CARBOXYLATION OF THE COVALENTLY  
 ATTACHED BIOTIN IN THE FIRST STEP AND THE TRANSFER OF THE  
 CARBOXYL GROUP TO PYRUVATE IN THE SECOND.  
 CC -1- CATALYTIC ACTIVITY: ATP + PYRUVATE + HCO(3) (-) = ADP +  
 ORTHOPHOSPHATE + OXALOACETATE.  
 CC -1- PATHWAY: GLUCONEOGENESIS.  
 CC -1- SUBUNIT: HOMOTETRAMER.  
 CC -1- COFACTOR: BIOTIN, AND ZINC.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- INDUCTION: BY GLUCOSE.  
 CC -1- SIMILARITY: WITH OTHER BIOTIN CARBOXYLASES, LIPONAMIDE  
 TRANSFERASES AND CARBAMYL PHOSPHATE SYNTHETASES.  
 DR EMBL; X59897; SCYB2G.  
 DR EMBL; X36087; SCYBR218C.  
 DR PIR; S46094; S46094.  
 DR PROSITE; PS00186; BIOTIN.  
 DR PROSITE; PS00866; CPSASE 1.  
 DR PROSITE; PS00867; CPSASE 2.  
 KM LIGASE; MULTIFUNCTIONAL ENZYME; BIOTIN; GLUCONEOGENESIS; ZINC;  
 KW MULTIGENE FAMILY.  
 FT NP BIND 183 188 ATP (POTENTIAL).  
 FT ACT SITE 313 313 BY SIMILARITY.  
 FT BINDING 1136 1136 BIOTIN (BY SIMILARITY).  
 FT SIMILAR 161 331 CARBAMOYL PHOSPHATE SYNTHETASES.  
 FT SIMILAR 331 471 WITH OTHER BIOTIN CARBOXYLASES.  
 FT SIMILAR 1087 1180 WITH OTHER BIOTIN CARRIER PROTEINS AND  
 WITH LIPONAMIDE ACETYLTTRANSFERASE.  
 FT CONFLICT 15 15 S -> C (IN REF. 1).  
 FT CONFLICT 132 132 D -> E (IN REF. 1).  
 FT CONFLICT 238 238 N -> K (IN REF. 1).  
 FT CONFLICT 268 268 L -> F (IN REF. 1).  
 FT CONFLICT 546 546 S -> C (IN REF. 1).  
 FT CONFLICT 642 642 N -> T (IN REF. 1).  
 FT CONFLICT 771 773 GTA -> STR (IN REF. 1).  
 FT CONFLICT 831 831 W -> R (IN REF. 1).  
 FT CONFLICT 839 839 S -> P (IN REF. 1).  
 FT CONFLICT 1001 1001 Y -> N (IN REF. 1).  
 FT CONFLICT 1155 1155 K -> R (IN REF. 1).  
 FT CONFLICT 1178 1178 Q -> P (IN REF. 1).  
 FT CONFLICT 1180 1180 K -> KYIFTR (IN REF. 1).  
 SQ SEQUENCE 1180 AA; 130166 MW; 7042611 CN;  
 Initial Score = 22 Optimized Score = 68 Significance = 7.74  
 Residue Identity = 23% Matches = 91 Mismatches = 237  
 Gaps = 65 Conservative Substitutions = 0  
 X 10 20 30 40 50  
 MELTELLIVMLLTARLLTSSPAPACDLRWL-----KLIDSHVLSRLSQCEVHPLE  
 RIRCEKTNIPILLITLTNVEFTEGR-YWTFIDDPQLQGWSSONRAKXLIHLIDIAVNGSISKGQIGLH  
 440 X 450 460 470 480 490 500



TPVLLPAVDVSLGEWKTQMEETKADILIGAVTLLEGMAARGOLGPTCLSLGQSGVRLT-----LGA  
 KKSPPSPVPHLDACQANVINTKSPSGMROVLE-----KGPSEKAVQROFNG--TLMDTWROA  
 510 520 530 540 550 560  
 130 140 150 160 170 180  
 LOSILGT-----QLPQGRRTAHKDPNAIFLSFQHLRGKV--RFLMLVGGSTLCVRRAPPTTAVPSRSTLV  
 HOSLIATVRTHDLATLPTATTAHALAGALECGCATFPVNAERFLHEDPWELRLR-----SLVPNTLPFOA  
 570 580 590 600 610 620 630  
 190 200 210 220 230 240  
 LTNELPRTSGL---LETNFTASARTG-----SGLKWOQ---GFR--KIPGLNQT-SRSD-QLP  
 LIRGANGVAYSLSLDPNAIDHFVKQAKONGVDIFRVPALNDLEQAKGVNAVAKAGVYEAIVCISGMLDP  
 640 650 660 670 680 690 700  
 250 260 270 280 290 300 310  
 G-YLNRHIELNGRTGLFPGPSRRRLGAPDISGTS--SDTGLPMLAGYSPSPHPPTGQYTLFLPPT  
 CKKRLDYLLEVEKIVOMG--THILGIKD-MAGTKPAKALLIGSLKTRYPDPIH-----VISHSAGI  
 710 720 730 740 750 760 770  
 320 330 340 350 X  
 LPTPVQLPPLPDPASAPTPPTSPILNTSYTHSONL--SOEG  
 AVASTACALAGADVVDVAIVNSMGL--TSOPSNALLASLEGNIDTGIVEH  
 780 790 800 810 X 820  
 7. US-08-223-263-1 (1-353)  
 FOR4\_MOUSE FORMIN 4.  
 ID FOR4\_MOUSE STANDARD; PRT; 1206 AA.  
 AC Q05859;  
 DT 01-JUN-1994 (REL. 29, CREATED)  
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)  
 DE FORMIN 4.  
 GN LD.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUDAROTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 EUHERIA; RODENTIA.  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC TISSUE=EMBRYO;  
 RM 92112033  
 RA GRUSBY-JACKSON L., KOO A., LEDER P.;  
 RL GENES DEV. 6:29-37 (1992).  
 CC -1- FUNCTION: IS IMPORTANT IN THE MORPHOGENESIS OF LIMB  
 AND MAY HAVE A FUNCTION IN DIFFERENTIATED CELLS OR  
 BE INVOLVED IN MAINTAINING SPECIFIC DIFFERENTIATED STATES.  
 CC -1- TISSUE SPECIFICITY: IT IS FOUND THROUGHOUT THE EMBRYO BUT  
 HAS A FUNCTIONAL ROLE ONLY IN THE LIVER AND LIMB.  
 CC -1- DEVELOPMENTAL STAGE: THIS IS THE ISOFORM FOUND IN THE APICAL  
 ECTODERMAL RIDGE AND THE MESENCHYMAL COMPARTMENT OF THE DEVELOPING  
 LIMB BUD.  
 CC -1- ALTERNATIVE PRODUCTS: MANY DIFFERENT FORMS OF THIS PROTEIN ARE  
 PRODUCED BY ALTERNATIVE SPLICING OF THE ID GENE. A VARIATION IN  
 SPLICING IS SEEN AMONG DIFFERENT TISSUES AND DIFFERENT SIZE

CC TRANSCRIPTS EXIST WITHIN ANY ONE TISSUE. THIS IS ISOFORM 4 AND IS  
 DIFFERENT IN ITS N-TERMINAL TO THE OTHER MOUSE ISOFORMS SO FAR  
 CC DETERMINED.  
 DR EMBL; X62379; MAFOR.  
 DR PIR; S24407; S24407.  
 KW NUCLEAR PROTEIN; DEVELOPMENTAL PROTEIN; ALTERNATIVE SPLICING.  
 FT DOMAIN 635 638 POLY-SER.  
 FT DOMAIN 644 744 PRO-RICH.  
 FT DOMAIN 751 755 POLY-SER.  
 SQ SEQUENCE 1206 AA; 133464 MW; 7742033 CN;  
 Initial Score = 22 Optimized Score = 67 Significance = 7.74  
 Residue Identity = 23% Matches = 95 Mismatches = 220  
 Gaps = 95 Conservative Substitutions = 0  
 X  
 MELTE-----LLVVMLLTARLTLSSPAPACDLRLVLSKILNDSHYL--HSRLSCPE  
 EQLSQLINIDMPRTOKRADEPFHGADEMGYSTDOESHRGP-BDAHVGQGVARTPEPALAFALTRP-  
 330 X 360 370 380 390 400 410  
 60 70 80 90 100 110 120  
 VHPLETPVLLPAVDVS-LGEWKTQMEETKADILIGAVTLLEGMAARGOLGPTCLSLGQSGVRLTIG  
 -----PKKSTADTSELEALKRKMKHEK--ESLRAY-----FRSKSRPADSPSDKSPDQPTDODRTG  
 420 430 440 450 460 470  
 130 140 150 160 170 180  
 ALQ-----SLGTQLPPQGRRTAHKDPNAIFLSFQHLR-GVRRFLMIVGSGSTLCVRRAPPTTA-VPSRTS  
 RLQAVWPPEKKTMDTEEKVGLKYTEAFYOAI-----LHKREHEELETLOAOFELKTFHIGEHVLNAR--  
 480 490 500 510 520 530  
 190 200 210 220 230 240  
 LVLTINE-----LPNRTSGLE-TNFTASARTGS---GLKWOQGR--AKIPGLNQTSRSDQIP  
 ---LEEALENLKQOLEKRRGCEEMDVOCISTDDSPAFNVGCIQDRETFLLKPCDAESKATRS-SQL-  
 540 550 560 570 580 590 600  
 250 260 270 280 290 300  
 GYLNRHIELNGRTGLFPGPSRRRLGAP-DISSGTS---DTGSLP-----PVLQPGYSPSPHPPTGQYTL  
 ----VPKRLTSLTQLSPSKSDKDIHAPFOTRECTSSSOOKISPAAPPTPPLPPLPLPPPPPLPG---L  
 610 620 630 640 650 660  
 310 320 330 340 350 X  
 FLIPPTLPVVV--QLHPILDPSPAPPTPTS-----PLNTSYTHSONLSOEG  
 GRLPAPPLPVPVCPVSPPPPPPPPTPVPVPSDGPPTPPPPPPPLN-----VIALPNSGGPPPPPPP  
 670 680 690 700 710 720 X 730  
 8. US-08-223-263-1 (1-353)  
 FORM\_MOUSE FORMIN.  
 ID FORM\_MOUSE STANDARD; PRT; 1468 AA.  
 AC Q05860;  
 DT 01-JUN-1994 (REL. 29, CREATED)  
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)  
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)

```

DE FORMIN
OS ID.
OS MDS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
RN EUTHERIA; RODENTIA.
  (1)
RM SEQUENCE FROM N.A.
RM TISSUE=KIDNEY, AND TESTIS;
RA 90363291
RA WOJCHIK R.P., MAAS R.L., ZELLER R., VOGT T.F., LEDER P.;
RL NATURE 346:850-853(1990).
  -1- FUNCTION: IS IMPORTANT FOR THE MORPHOGENESIS OF LIMB AND KIDNEY
CC AND MAY HAVE A FUNCTION IN DIFFERENTIATED CELLS OR MAY BE
CC INVOLVED IN MAINTAINING SPECIFIC DIFFERENTIATED STATES.
  -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: IT IS PRESENT IN THE ADULT KIDNEY, TESTIS,
CC LIMB, OVARY, BRAIN, SMALL INTESTINE, SALIVARY GLAND AND HARDERIAN
CC GLAND. IT IS PRESENT THROUGHOUT THE EMBRYO.
  -1- DEVELOPMENTAL STAGE: IN THE DEVELOPING LIMB BUD, THE PROTEIN
CC IS EXPRESSED IN THE APICAL ECTODERMAL RIDGE AND THE MESenchYMAL
CC COMPARTMENT, PREDOMINANTLY IN THE POSTERIOR REGION. DURING
CC KIDNEY MORPHOGENESIS, EXPRESSION IS INITIALLY RESTRICTED TO
CC THE EPITHELIAL COMPARTMENT OF THE PRONEPHROS AND MESONEPHROS.
  -1- ALTERNATIVE PRODUCTS: MANY DIFFERENT FORMS OF THIS PROTEIN
CC ARE PRODUCED BY ALTERNATIVE SPLICING OF THE ID GENE. A
CC VARIATION IN SPLICING IS SEEN AMONG DIFFERENT TISSUES AND
CC DIFFERENT SIZE TRANSCRIPTS EXIST WITHIN ANY ONE TISSUE.
DR EMBL; X53599; MAFDP.
DR PIR; S11513; S11513.
KW NUCLEAR PROTEIN; DEVELOPMENTAL PROTEIN; ALTERNATIVE SPLICING.
FT DOMAIN 298 301 POLY-SER.
FT DOMAIN 861 864 POLY-SER.
FT DOMAIN 869 970 PRO-RICH.
FT DOMAIN 977 981 POLY-SER.
SQ SEQUENCE 1468 AA; 163809 MW; 11493196 CN;

Initial Score = 22 Optimized Score = 95 Significance = 7.74
Residue Identity = 23% Matches = 97 Mismatches = 224
Gaps = 86 Conservative Substitutions = 0

      X      10      20      30      40      50
MEITELLIVMLLTARLT-LSPAPACDLRLSKILRDSHYL-----HSRLSQCEVHPILP
      ||      ||      ||      ||      ||      ||
SADTLEPSSITVTEKASPTSLASQTWIVSEASRKGIPTAPPOHDLPGIASSEGPC-DNFKEQ
570      580      590      600      610      620      630

      60      70      80      90      100      110      120
TPVILPAVDFSLGEKWTQMEETKADIIIGAVTLLLEGVMARGLGFTCLSLAQ-LSGYVRLILGALQ-----
      ||      ||      ||      ||      ||      ||      ||
TANDLEPKND--GGVWVPGVRACPPCPFL-----LHEEKRTSRSEL--YIDLPDQSPTEODDRTPGLGLAV
640      650      660      670      680      690      700

      130      140      150      160      170      180      190
-----SLIGQLPQPGKRTTHKDPNALIFSLFOHLIR-GKVRIMLVGSLTCVRRAPPTTA-VPSTSLVTLTL
      ||      ||      ||      ||      ||      ||      ||
WPPPKKQDVEEKVGLKYTEAEVQAL-----LHKREKHEEITELQALFELKTHIRGIEALVTAR-----L
710      720      730      740      750      760

NE-----LKNRTISGLLE-TNFTASARTGS---GLTKWQOGFR---AKIPGLINQTSRLDIOIPGYNIR

```

```

EALINIKQQLKKRECEGEEMRDVCI1STDDCSPPAFRNVCIQTDREFFLKPCDAE5KATRS-SQ1-----V
770      780      790      800      810      820      830

250      260      270      280      290      300      310
IHLINGTRKGLFPGSRRTLGAP-DISGGRS-----DTGSLP-----PNLGGYSPSTHPTGQYTLPLP
      1      2      3      4      5      6      7      8      9      10     11     12
PKKTLISLTQSPKSDSKDIIHAPQRRECGSSSQQKISPPAPRPRLPRLPLRPPLRPPLPC---LcglRP
      400      450      500      550      600      650      700      750      800      850      900

320      330      340      350 X
TLTPPV--QLHPLLPDSAPTPTIS-----PLNTSYTHSQNLSQEG
      1      2      3      4      5      6      7      8      9      10     11     12
APRIIPVCGVSPPPPPPPPPPTTVPRSDCGPPPPPPPPPLRN-----VLALPNSGGCGPPPPPPPP
      900      910      920      930      940      950

```

	9.	US-08-223-263-1	(1-353)	
EPO_RAT		ERYTHROPOIETIN PRECURSOR.		
ID	EPO_RAT	STANDARD;	PRT;	192 AA.
AC	P29676f;			
DT	01-APR-1993	(REL. 25, CREATED)		
DT	01-APR-1993	(REL. 25, LAST SEQUENCE UPDATE)		
DT	01-FEB-1994	(REL. 28, LAST ANNOTATION UPDATE)		
DE	ERYTHROPOIETIN PRECURSOR.			
GN	EPO.			
OS	RATTUS NORVEGICUS (RAT).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;			
OC	EUTHERIA; RODENTIA.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MISTAR; TISSUE=KIDNEY;			
RM	93042015			
RA	NAGAO M., SUGA H., OKANO M., MASUDA S., NARITA H., IKURA K.,			
RA	SASAKI R.;			
RL	BIOCHIM. BIOPHYS. ACTA 1171:99-102(1992).			
CC	-  FUNCTION: ERYTHROPOIETIN IS THE PRINCIPAL HORMONE INVOLVED IN THE REGULATION OF ERYTHROCYTE DIFFERENTIATION AND THE MAINTENANCE OF A PHYSIOLOGICAL LEVEL OF CIRCULATING ERYTHROCYTE MASS.			
CC	-  TISSUE SPECIFICITY: PRODUCED BY KIDNEY OR LIVER OF ADULT MAMMALS AND BY LIVER OF FETAL OR NEONATAL MAMMALS.			
CC	-  SUBCELLULAR LOCATION: SECRETED.			
DR	EMBL; D10763; RNMPD.			
DR	PIR; S28148; S28148.			
DR	PROSITE; PS00817; EPO.			
KW	ERYTHROCYTE MATURATION; GLYCOPROTEIN; HORMONE; SIGNAL.			
FT	SIGNAL	1	26	BY SIMILARITY.
FT	CHAIN	27	192	ERYTHROPOIETIN.
FT	DISTIFTD	33	187	BY SIMILARITY.
FT	CAROHYD	50	50	BY SIMILARITY.
FT	CAROHYD	64	64	BY SIMILARITY.
FT	CAROHYD	109	109	BY SIMILARITY.
SQ	SEQUENCE	192 AA;	21286 MW;	179242 CN;
Initial Score =	21	Optimized Score =	44	Significance = 7.25
Residue Identity =	27%	Matches =	53	Mismatches = 118
Caps =	24	Conservative Substitutions =		0

x	10	20	30	40	50	60
---	----	----	----	----	----	----



250 260 270 280 290 300 310  
 QIPGYLNRIHELLNCTRGCLFPGPSRRITGAPDISSGTSDTSLPNILOPGYSPSPHPPTGQYTLFPLPPTL  
 PDGCTFNNDLITQVAKIKYFSPGPHGHDOQYVYTWALAFDPPWVKRPPVHKRP-PP-----PLPPSA  
 60 70 80 90 100 110  
 320 330 340 350 X  
 PTPVVOHLPLDPSAPTPTPTSPILNTS---YTHSQNLISQEG  
 P-----SIPLEPPLSTPPRSSIXPALTPSLGAKPKPQVLSDSGGLDILTED  
 120 130 140 150 160

12. US-08-223-263-1 (1-353)

GAG\_MSVMT GAG POLYPROTEIN (CONTAINS: CORE PROTEIN P15; INNER

ID GAG\_MSVMT STANDARD; PRT; 468 AA.

AC P32594;  
 DT 01-OCT-1993 (REL. 27, CREATED)  
 DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)  
 DE GAG POLYPROTEIN (CONTAINS: CORE PROTEIN P15; INNER COAT PROTEIN P12;  
 DE CORE SHELL PROTEIN P30).  
 GN GAG.  
 OS MOLONEY MURINE SARCOMA VIRUS (STRAIN TS110).  
 OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;  
 OC ONCOVIRINAE.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RM 92365121  
 RA HUAI L., CHIOCCA S.M., GILBRETH M.A., AINSWORTH J.R., BISHOP L.A.,  
 RA MURPHY E.C., JR.;  
 RL J. VIROL. 66:5329-5337(1992).  
 CC -1- P15: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
 CC -1- THIS PROTEIN IS PROBABLY TRANSLATED AS A GAG-MOS POLYPROTEIN.  
 DR EMBL; M96854; MMSAAX.  
 DR EMBL; S42321; S42321.  
 DR PIR; A42745; FOMVNU.  
 KW COAT PROTEIN; CORE PROTEIN; POLYPROTEIN; MYRISTYLATION.  
 FT CHAIN 2 131  
 FT CHAIN 132 215  
 FT CHAIN 216 468  
 FT LIPID 2 2  
 FT MYRISTATE (BY SIMILARITY).  
 SQ SEQUENCE 468 AA; 52681 MW; 1077859 CN;

Initial Score = 21 Optimized Score = 35 Significance = 7.25  
 Residue Identity = 22% Matches = 40 Mismatches = 120  
 Gaps = 15 Conservative Substitutions = 0

170 180 190 200 210 220 230 240  
 TLGVRRAPRTAVPSRTSLVTLNLELNPRTSGLETFNTASARTGSLKWKQGFRAKIPGLNLTSSLD  
 X MCGVTTPSLTL-----DHKQVERIAHNSQSDVYKRRRWVTFCSAEWF-----TFNVGW  
 10 20 30 40 50

320 330 340 350 X  
 PT-PVVOHLPLDPSAPTPTPTSPILNTSYTHSQNLISQEG  
 STPPOSSILYPAL-TFS-LGAKPKPQVLSDSGGLDILTEDPPPYDPPRP  
 130 140 150 160 170

13. US-08-223-263-1 (1-353)

OCH1\_YEAST PROBABLE ALPHA-1,6-MANNOSYLTRANSFERASE (EC 2.4.1.-

ID OCH1\_YEAST STANDARD; PRT; 480 AA.

AC P31753;  
 DT 01-JUL-1993 (REL. 26, CREATED)  
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)  
 DE PROBABLE ALPHA-1,6-MANNOSYLTRANSFERASE (EC 2.4.1.-).  
 GN OCH1.  
 OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).  
 OC EUCARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EMF-2C;  
 RM 92331603  
 RA NAKAYAMA K.-I., NAGASU T., SHIMAWA Y.-I., KURUMITSU J.-R., JIGAMI Y.;  
 RL EMBO J. 11:2511-2519(1992).  
 CC -1- FUNCTION: INVOLVED IN OUTER CHAIN ELONGATION OF ASPARAGINE-LINKED  
 CC OLIGOSACCHARIDES OF THE TYPE MAN(9)G1CNA(2).  
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ANCHORED TO  
 CC THE ER AND/OR THE GOLGI MEMBRANES.  
 DR EMBL; S40019; S40019.  
 DR EMBL; D11095; SCCH1.  
 DR PIR; S22701; S22701.  
 KW TRANSFERASE; GLYCOSYLTRANSFERASE; TRANSMEMBRANE; SIGNAL-ANCHOR;  
 KW GLYCOPROTEIN.  
 FT DOMAIN 1 15  
 FT DOMAIN 16 30  
 FT DOMAIN 31 480  
 FT CARBOHYD 203 203  
 FT CARBOHYD 281 281  
 FT CARBOHYD 341 341  
 FT CARBOHYD 393 393  
 SQ SEQUENCE 480 AA; 55155 MW; 1228766 CN;

Initial Score = 21 Optimized Score = 56 Significance = 7.25  
 Residue Identity = 20% Matches = 73 Mismatches = 236  
 Gaps = 43 Conservative Substitutions = 0

20 30 40 50 60 70 80  
 TLSSPAPACDLRLVLSKI--LRDS--HYLHSLRSLQCEVHPILPVLPAVDLSLEWKTOMEET--KAQDI  
 X MSKRSLHILATRKSKTIYVYVLLIYSLITFLHSKRLLSQFYFSKDDFQYTLPTTSHSDI  
 10 20 30 40 50 60

[illegible]

14.	US-08-223-263-1 (1-353)	
	GAG_MSVMO	GAG POLYPEPTIDE R65 (CONTAINS: CORE PROTEIN P15; I
AD	GAG_MSVMO	STANDARD; PRT; 538 AA.
IC	P03334;	
DT	21-JUL-1986 (REL. 01, CREATED)	
DT	21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)	
DT	01-MAR-1992 (REL. 21, LAST ANNOTATION UPDATE)	
DE	GAG POLYPEPTIDE R65 (CONTAINS: CORE PROTEIN P15; INNER COAT PROTEIN P12; CORE SHELL PROTEIN P30; NUCLEOPROTEIN P10).	
GN	GAG.	
OS	MOLONEY MURINE SARCOMA VIRUS.	
OC	VIRAL; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; RETROVIRIDAE;	
OC	ONCOVIRINAE.	
RA	[1]	
RA	PROVIRUS, SEQUENCE FROM N.A.	
RA	82039539	
RA	REDDY E.P., SMITH M.J., AARONSON S.A.;	
RL	SCIENCE 214:445-450 (1981).	
RN	[2]	
RP	CLONE 124 CIRCULAR, SEQUENCE FROM N.A.	
RM	82115347	
RA	VAN BEVEREN C., VAN STRAATEN F., GALLISHAW J.A., VERA I.M.;	
RL	CELL 27:97-108 (1981).	
CC	-1 PPM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.	
DR	PIR; A03932; FOMM.	
KW	COAT PROTEIN; CORE PROTEIN; NUCLEOPROTEIN; POLYPEPTIDE; MYSTIFICATION	
FT	CHAIN 2 131	
FT	CHAIN 132 215	
FT	CHAIN 216 478	
FT	CHAIN 479 534	
FT	CHAIN	NUCLEOPROTEIN P10.

FT	LIPID	2	2	MYRISTATE.
FT	VARIANT	519	519	R -> K (IN CLONE 124) .
ISO	SEQUENCE	538 AA;	61209 MW;	1333981 CN;

  

initial Score	=	21	Optimized Score	=	36	Significance	=	7.25
Residue Identity	=	2%	Matches	=	43	Mismatches	=	114
aps	=	21	Conservative Substitutions	=			=	0

  

170	180	190	200	210	220	230
TLCVRAPTTAAVPSRSLVLTINELPNRSGLLETNTFASRTTGSGLK---	WQGFPAKIPGLNQTSE					
MGQVTPPLSLT---	DHMKDVERLAHN-----	QSDVYKKRMWTFCSAEWF-----	TFN			
X	10	20	30	40		

  

240	250	260	270	280	290	300	310
SLLDQIGYLIINLHELLNGTIGLPGFSRRRTLAGPDISSGTSIDTSGSLPNLPGYSPSPHTHPPGQYTLPLPL							
VGMPSDGFNFNDLITQVRIKVFSPGHGHPDOVPYIVTWEALAFDPWPVWFVHPKRPPLLPASLAPLE							
50	60	70	80	90	100	110	

  

320	330	340	350	X	
PLTPT-PVVOQLHPLLPDSAPPTPTPSPLINTSYTHSONLSSEG					
PLPLSTPQSSSLYPAL-TPS--IGAKPKPQYLSNSGGELIDLITEDPPPYRDRPP					
120	130	140	150	160	170





maryh@stic

stdin

NeWSprinter20

Tue Apr 11 13:29:56 1995

NeWSprint 2.5 Rev B

Openwin library 3

NeWSprint interpreter 210.0

NeWSprint 2.5